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 η-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 PrivUnit2 (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 η 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 η-local differential privacy (η-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).

∗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.
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 $\varepsilon$-LDP constraints, Chen et al. (2020) presents minimax order-optimal mechanisms for frequency and mean estimation that require only $\varepsilon$ bits (independent of the underlying dimensionality of the problem) by using shared randomness. 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 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 $\varepsilon$-LDP while only using on the order of $\varepsilon$-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 $\varepsilon$-LDP mechanism in a near-lossless fashion using only on the order of $\varepsilon$-bits of communication (see Theorem 3.1). We also prove that the resulting compressed mechanism is $2\varepsilon$-LDP (see Theorem 3.2). Thus, to achieve $\varepsilon$-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 $\varepsilon$-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 $\delta$ (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 $\varepsilon$-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 $\varepsilon$ 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 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, via a variety of datasets, that MMRC achieves accuracy comparable to PrivUnit$_2$ and Subset Selection (see Section 4.1 and 5.1) while only using about $\varepsilon$ bits. See the Appendix for additional results and experiments.

1.2 Related Work

A number of recent works have examined approaches for compressing LDP schemes in the presence of shared randomness. When $\varepsilon = O(1)$, Bassily and Smith (2015) show that a single bit is enough to simulate any LDP randomizer with (almost) no impact on its utility, and Bassily et al. (2017), Bun et al. (2019), Acharya and Sun (2019) propose 1-bit order-optimal schemes for frequency estimation. The recent work of Chen et al. (2020) generalizes these methods to arbitrary $\varepsilon$’s and provides order-optimal schemes for both frequency and mean estimation that only use on the order of $\varepsilon$ bits. However, all these methods are only order-optimal and cannot achieve the exact accuracy of the best known schemes: PrivUnit$_2$ (Bhowmick et al., 2018) for

---

1We assume that the encoder and decoder can depend on a random quantity that both the server and user have access to. See Section 2 for details.
mean estimation, and Subset Selection (Ye and Barg, 2018) \(^2\) for frequency estimation. We show how one can achieve the accuracy of these schemes with \(\varepsilon\) bits of communication.

In the absence of shared randomness, the works of Girgis et al. (2021b), Girgis et al. (2021a), Chen et al. (2020) provide order-optimal mechanisms for frequency and mean estimation but their mechanisms do not achieve the best known accuracy. The recent work of Feldman and Talwar (2021) presents an approach for compressing \(\varepsilon\)-LDP schemes in a lossless fashion using a pseudorandom generator (PRG). Their approach can compress Subset Selection to \(O(\ln d)\) bits and PrivUnit\(_2\) to \(O(\varepsilon + \ln d)\) bits, where \(d\) is the dimension of the underlying problem. Nevertheless, their approach is designed to work without shared randomness, therefore requiring more bits than needed if shared randomness is available as in our work.

Unlike previous works, our technique of compressing generic LDP schemes relies on Minimal Random Coding (MRC), which is originally 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 show 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 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, the problems of mean and frequency estimation under LDP constraints, two canonical problems in distributed learning and analytics, have been widely studied in the literature (Duchi et al., 2013; Nguyen 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

Locally Differentially Private (LDP). Suppose \(x \in \mathcal{X}\) is some data that must remain private. A privatization mechanism \(q\) is a randomized mapping that maps \(x \in \mathcal{X}\) to \(z \in \mathcal{Z}\) with probability \(q(z|x)\) where \(\mathcal{Z}\) can be arbitrary. Further, \(q\) is \((\varepsilon, \delta)\)-LDP if

\[
\forall x, x' \in \mathcal{X}, z \in \mathcal{Z}, q(z|x) \leq \exp(\varepsilon)q(z|x') + \delta.
\]

When \(\delta = 0\), the mechanism is \(\varepsilon\)-LDP. Here, we focus on \(\varepsilon\)-LDP mechanisms where \(\varepsilon \geq 1\).

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

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

We allow the estimator of \(x\) at the server to implicitly depend on \(u\). However, for simplicity, we suppress the dependence on \(u\) in our notation. In practice, such shared randomness can be achieved via downlink communication, that is, the sever generates \(u\) (e.g., a random seed) and communicates it to the user.

\(^2\)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.
**PrivUnit2.** The PrivUnit2 mechanism \(q^{pu}\), proposed by Bhowmick et al. (2018), is an \(\varepsilon\)-LDP sampling scheme when the input alphabet \(\mathcal{X}\) is the \(d\)-dimensional unit \(\ell_2\) sphere \(S^{d-1}\). Formally, given a vector \(x \in S^{d-1}\), PrivUnit2 draws a random vector \(z\) from a spherical cap \(\{z \in S^{d-1} | \langle z, x \rangle \geq \gamma \}\) with probability \(p_0\) or from its complement \(\{z \in S^{d-1} | \langle z, x \rangle < \gamma \}\) with probability \(1 - p_0\), where \(\gamma \in [0, 1]\) and \(p_0 \geq 1/2\) are parameters (depending on \(\varepsilon\) and \(d\)) that trade accuracy and privacy. The estimator of the PrivUnit2 mechanism (denoted by \(\hat{x}^{pu}\)) is defined as \(z/m_{pu}\) where

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

with \(\alpha = (d - 1)/2\), \(\tau = (1 + \gamma)/2\), and \(B(x; \alpha, \beta)\) denoting the incomplete beta function. The estimator \(\hat{x}^{pu}\) is unbiased, \(E[\hat{x}^{pu}|x] = x\) and has order-optimal utility, \(E[||\hat{x}^{pu} - x||_2^2] = \Theta\left(\frac{d}{\min(\varepsilon, (\varepsilon^2 - 1)^2 \tau^2 d^2)}\right)\). See Appendix C for more details on PrivUnit2.

**Subset Selection.** Subset Selection (Ye and Barg, 2018) is an \(\varepsilon\)-LDP sampling scheme where the input alphabet \(\mathcal{X}\) is \([d] = \{1, \cdots, d\}\). Without loss of generality, let \(\mathcal{X} := \{e_1, e_2, \cdots, e_d\}\), 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 \(s := \lceil \frac{d}{1 + \varepsilon^2} \rceil\), i.e.,

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

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

\[
q^{ss}(z|x) := \begin{cases} 
\frac{e^\varepsilon}{(s-1)e^\varepsilon + (d-s)} & \text{if } z \in Z_x \\
\frac{1}{(s-1)e^\varepsilon + (d-s)} & \text{if } z \in \mathcal{Z} \setminus Z_x
\end{cases}
\]

where \(Z_x = \{ z = (z^{(1)}, \cdots, z^{(d)}) \in \mathcal{Z} : \langle x, z \rangle = 1 \}\) is the set of elements in \(\mathcal{Z}\) with 1 in the \(x^{th}\) location. The estimator of the Subset Selection mechanism (denoted by \(\hat{x}^{ss}\)) is defined as \((z - b_{ss})/m_{ss}\) where

\[
m_{ss} := \frac{s(d - s)(e^\varepsilon - 1)}{(d - 1)(s(e^\varepsilon - 1) + d)} \quad b_{ss} := \frac{s((s - 1)e^\varepsilon + (d - s))}{(d - 1)(s(e^\varepsilon - 1) + d)}.
\]

The estimator \(\hat{x}^{ss}\) is unbiased, \(E[\hat{x}^{ss}|x] = x\) and has order-optimal utility, \(E[||\hat{x}^{ss} - x||_2^2] = \Theta\left(\frac{d}{\min(e^\varepsilon, (e^\varepsilon - 1)^2 \tau^2 d^2)}\right)\). See Appendix F for more details on Subset Selection.

### 3 Main Results

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

---

\(^3\)The family of cap-based mechanisms includes PrivUnit2 and Subset Selection. See Definition 3.1.
Consider an \(\varepsilon\)-LDP mechanism \(q(\cdot|x)\) that we wish to compress. Under MRC, first, a number of candidates \(z_1, \cdots, z_N\) are drawn from a fixed reference distribution \(p(\cdot)\) (known to both the user and 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 \(\pi_{\text{mrc}}(\cdot)\) such that \(z_K \sim q(\cdot|x)\) approximately. The distribution \(\pi_{\text{mrc}}\) is such that, \(\forall k \in [N], \pi_{\text{mrc}}(k) \propto w(k)\) where \(w(k) := q(z_k|x)/p(z_k)\) are the importance weights\(^4\) (see Algorithm 1). To communicate the index \(K\) of MRC, \(\log N\) bits are required.

\[\pi_{\text{mrc}}(\cdot) := \frac{q(z_k|x)}{p(z_k)^{\varepsilon}} \frac{\sum_k w(k)}{\sum_k w(z_k)},\]

Let \(q_{\text{mrc}}^{\text{mrc}}\) denote the distribution of \(z_K\) where \(K \sim \pi_{\text{mrc}}(\cdot)\). The following theorem shows that when the number of candidates is exponential in \(\varepsilon\), samples drawn from \(q_{\text{mrc}}\) will be similar to samples drawn from \(q(\cdot|x)\) in terms of \(\ell_2\) error. In other words, \(q_{\text{mrc}}\) can compress \(q(\cdot|x)\) to the order of \(\varepsilon\) bits of communication as well as simulate it in a near-lossless fashion. A proof can be found in Appendix A.1.

**Theorem 3.1 (Utility of MRC).** Consider any input alphabet \(X\), output alphabet \(Z\), data \(x \in X\), and \(\varepsilon\)-LDP mechanism \(q(\cdot|x)\). Consider any reference distribution \(p(\cdot)\) such that \(\|\ln(q(z|x)/p(z))\|_\infty \leq \varepsilon \forall x \in X, z \in Z\).\(^5\) Let the number of candidates be \(N = 2^{(\log \varepsilon + 4)c}\) for some constant \(c \geq 0\). Then, \(q_{\text{mrc}}\) is such that

\[\left| \mathbb{E}_{q_{\text{mrc}}}(\|z - x\|^2) - \mathbb{E}_q(\|z - x\|^2) \right| \leq \frac{2\alpha \sqrt{\mathbb{E}_q(\|z - x\|^4)}}{1 - \alpha}\]

holds with probability at least \(1 - 2\alpha\), where

\[\alpha = \sqrt{2^{-\varepsilon^2} + 2^{-\varepsilon^2}/\log \varepsilon + 1}.\]

In general, \(\mathbb{E}_q(\|z - x\|^4)\) in (5) can be well-controlled. See Remark A.1 in Appendix A.1 for more details.

In the next Theorem, we show that \(\pi_{\text{mrc}}\) is \(2\varepsilon\)-LDP. Hence, the compressed mechanism \(q_{\text{mrc}}\) is \(2\varepsilon\)-LDP.

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

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

\[\frac{\pi_{\text{mrc}}(k)}{\pi_{\text{mrc}}'(k)} = \frac{q(z_k|x)}{q(z_k'|x')} \frac{\sum_k q(z_k|x')}{\sum_k q(z_k'|x')}.\]

\(^4\)We suppress dependence of \(\pi_{\text{mrc}}\) & \(w\) on \(x\) for simplicity.

\(^5\)Note that this condition holds for many reference distributions \(p(\cdot)\). For example, one can simple choose \(p(\cdot) = q(\cdot|x)\) for some \(x' \in X\).
In the following Theorem, we show that $\pi_{\text{src}}$ is $(\varepsilon + \varepsilon_0, \delta)$-LDP implying that the compressed mechanism $q_{\text{src}}$ is $(\varepsilon + \varepsilon_0, \delta)$-LDP where $\varepsilon_0 > 0$ and $\delta \leq 1$ are free parameters. This Theorem can be viewed complimentary to Theorem 3.2 where a stronger privacy parameter can be achieved (i.e., $\varepsilon + \varepsilon_0$ which can get arbitrarily close to $\varepsilon$ as opposed to $2\varepsilon$) albeit at the cost of trading pure privacy for approximate privacy. A proof is provided in Appendix A.2.2.

**Theorem 3.3 (Approximate DP guarantee of MRC).** Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $x \in \mathcal{X}$, and $\varepsilon$-LDP mechanism $q(\cdot|x)$. Consider any reference distribution $p(\cdot)$ such that $\left|\ln(q(z|x)/p(z))\right| \leq \varepsilon \forall x \in \mathcal{X}, z \in \mathcal{Z}$.\(^6\) Let $c_0 \geq 0$ be some constant and let the number of candidates $N = \exp(2\varepsilon + 2c_0)$. Then, for any $\delta \leq 1$, $\pi_{\text{src}}(\cdot)$ obtained from Algorithm 1 is $(\varepsilon + \varepsilon_0, \delta)$-LDP mechanism where

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

### 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 $\varepsilon$ bits of communication, the resulting compressed mechanism is not exactly $\varepsilon$-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 $\varepsilon$-LDP mechanisms which we call *cap-based* mechanisms and propose a modification to MRC such that the resulting compressed mechanism is $\varepsilon$-LDP. Further, like MRC, MMRC can simulate the underlying $\varepsilon$-LDP mechanism in a near-lossless fashion while using only on the order of $\varepsilon$ bits.

We start with the definition of cap-based mechanism which is inspired from the structure of PrivUnit2 and Subset Selection.

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

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

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

In words, a cap-based $\varepsilon$-LDP mechanism samples uniformly either from $\text{Cap}_x$ or from $\mathcal{Z} \setminus \text{Cap}_x$ where $\text{Cap}_x \subseteq \mathcal{Z}$ is such that if $z$ is sampled uniformly from $\mathcal{Z}$, it will belong to $\text{Cap}_x$ with probability at least $c_2(\varepsilon, d)/2c_1(\varepsilon, d)$. It is easy to see that $q_{\text{src}}$ defined in (3) is a cap-based mechanism with $\text{Cap}_x = \mathcal{Z}_x$, $c_1(\varepsilon, d) = \frac{e^\varepsilon}{(\frac{e^\varepsilon}{1 - e^{-\varepsilon}} - 1)}$, and $c_2(\varepsilon, d) = \frac{1}{(\frac{e^\varepsilon}{1 - e^{-\varepsilon}} - 1)}$. See Appendix F where we evaluate $\mathbb{P}_{z \sim \text{Unif}(\mathcal{Z})}(z \in \mathcal{Z}_x)$ and show that it is at least $1/2e^\varepsilon$. See Appendix C where we show $q_{\text{src}}$ is a cap-based mechanism.

For a cap-based $\varepsilon$-LDP mechanism $q(z|x)$ and a uniform reference distribution $p(\cdot)$, the distribution $\pi_{\text{src}}$ obtained from Algorithm 1 takes a special form:

$$
\pi_{\text{src}}(k) = \begin{cases} 
    \frac{1}{N} \times \frac{c_1(\varepsilon, d)}{\theta c_1(\varepsilon, d) + (1 - \theta)c_2(\varepsilon, d)} & \text{if } z_k \in \text{Cap}_x \\
    \frac{1}{N} \times \frac{c_2(\varepsilon, d)}{\theta c_1(\varepsilon, d) + (1 - \theta)c_2(\varepsilon, d)} & \text{if } z_k \notin \text{Cap}_x
\end{cases}
$$

\(^6\)Note that this condition holds for many reference distributions $p(\cdot)$. For example, one can simple choose $p(\cdot) = q(\cdot|x^*)$ for some $x^* \in \mathcal{X}$.
where $\theta$ is the fraction of candidates inside the Cap$_x$, i.e., $\theta = \frac{1}{N} \sum_k I(z_k \in \text{Cap}_x)$. As is, $\pi^{\text{mrc}}$ in (7) is not necessarily $\varepsilon$-LDP because $\theta$ can be different for $x$ and $x'$. However, as $N \to \infty$, $\theta \to \mathbb{E}[\theta] = \mathbb{E} \sim \text{Unif}(\mathcal{Z}) (z \in \text{Cap}_x)$, which is not a function of $x$, implying that $\pi^{\text{mrc}}(k)/\pi^{\text{mrc}}(\cdot) \leq c_1(\varepsilon,d)/c_2(\varepsilon,d) \leq \exp(\varepsilon)^7$. This shows that $\pi^{\text{mrc}}$ is $\varepsilon$-LDP when $N \to \infty$. This motivates us to modify $\pi$ to $\pi^{\text{mrc}}$ such that $\pi^{\text{mrc}}$ is $\varepsilon$-LDP irrespective of $N$. Further, when $N$ is large enough, the modification is not by much i.e., a sample from $\pi^{\text{mrc}}$ is similar to a sample from $\pi$. 

To that end, define an upper threshold $t_u = \frac{1}{N} \sum_k c_1(\varepsilon,d) \pi^{\text{mrc}}(k)$ and a lower threshold $t_l = \frac{1}{N} \sum_k c_2(\varepsilon,d) \pi^{\text{mrc}}(k)$, and initialize $\pi^{\text{mrc}}$ to be equal to $\pi$. We want to modify $\pi^{\text{mrc}}$ so as to ensure:

\[ t_l \leq \pi^{\text{mrc}}(k) \leq t_u \forall k \in [N], \tag{8} \]

which, as argued above, guarantees $\varepsilon$-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 $\theta$. Next, we will look at 3 cases depending on the relationship between $\theta$ and $\mathbb{E}[\theta]$: (A) If $\theta = \mathbb{E}[\theta]$, then $\pi^{\text{mrc}}$ already satisfies (8); (B) If $\theta < \mathbb{E}[\theta]$, then only the upper threshold is violated and we set $\pi^{\text{mrc}}(k) = t_u \forall k : z_k \in \text{Cap}_x$ and re-normalize the remaining $\pi^{\text{mrc}}(k)$; (C) If $\theta > \mathbb{E}[\theta]$, then only the lower threshold is violated, we set $\pi^{\text{mrc}}(k) = t_l \forall k : z_k \notin \text{Cap}_x$ and re-normalize the remaining $\pi^{\text{mrc}}(k)$. It is easy to see that the re-normalization step does not violate (8). We provide pseudo-code to calculate $\pi^{\text{mrc}}$ in Algorithm 2.

**Algorithm 2: MMRC**

**Input:** $\varepsilon$-LDP cap-based mechanism $q(\cdot|x)$, the associated Cap$_x$, reference distribution $p(\cdot)$, number of candidates $N$, lower threshold $t_l$, upper threshold $t_u$  

\[
\pi^{\text{mrc}}(\cdot), \{z_1, \ldots, z_N\} \leftarrow \text{MMRC}(p(\cdot),q(\cdot|x),N)
\]

$\theta \leftarrow \frac{1}{N} \sum_k I(z_k \in \text{Cap}_x)$ \COMMENT{Compute the fraction of candidates inside the cap}

**Initialization:** $\pi^{\text{mrc}}(\cdot) \leftarrow \pi^{\text{mrc}}(\cdot)$

if $\max_k \pi^{\text{mrc}}(k) > t_u$ then  

\[
\pi^{\text{mrc}}(k) \leftarrow t_u, \forall k : z_k \in \text{Cap}_x \quad \pi^{\text{mrc}}(k) \leftarrow \frac{1 - N \theta u}{N(1 - \theta)}, \forall k : z_k \notin \text{Cap}_x
\]

else if $\min_k \pi^{\text{mrc}}(k) < t_l$ then  

\[
\pi^{\text{mrc}}(k) \leftarrow t_l, \forall k : z_k \notin \text{Cap}_x \quad \pi^{\text{mrc}}(k) \leftarrow \frac{1 - N(1 - \theta) t_l}{N \theta}, \forall k : z_k \in \text{Cap}_x
\]

**Output:** $\pi^{\text{mrc}}(\cdot), \{z_1, \ldots, z_N\}$

Let $q^{\text{mrc}}$ denote the distribution of $z_K$ where $K \sim \pi^{\text{mrc}}(\cdot)$. In the following Theorem, we show that $\pi^{\text{mrc}}$ is $\varepsilon$-LDP implying that the compressed mechanism $q^{\text{mrc}}$ is $\varepsilon$-LDP. The proof follows from (8) and can be found in Appendix B.1.

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

The following Theorem shows that, with number of candidates exponential in $\varepsilon$, samples drawn from $q^{\text{mrc}}$ will be similar to the samples drawn from $q$ in terms of $\ell_2$ error. A proof can be found in Appendix B.3.

\[\text{This follows from (1) and (6) because } q(\cdot|x) \text{ is } \varepsilon\text{-LDP}\]
**Theorem 3.5 (Utility of MMRC).** Consider any input alphabet $\mathcal{X}$, output alphabet $Z$, data $x \in \mathcal{X}$, and $\varepsilon$-LDP cap-based mechanism $q(\cdot|z)$. Let the reference distribution $p(\cdot)$ be the uniform distribution on $Z$. Let $N$ denote the number of candidates. Then, $q_{\text{mmrc}}$ is such that

$$
\mathbb{E}_{q_{\text{mmrc}}} \left[ \|z - x\|_2^2 \right] \leq \mathbb{E}_{q_{\text{mmrc}}} \left[ \|z - x\|_2^2 \right] + \sqrt{\frac{\rho(1+\varepsilon)}{2}} \max_{x,z} \|z - x\|_2^2
$$

where $\rho \in (0,1)$ is such that

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

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

In conjunction with Theorem 3.1, Theorem 3.5 implies that $q_{\text{mmrc}}$ can compress $q(\cdot|x)$ to the order of $\varepsilon$ bits of communication and simulate it in a near-lossless fashion. This is stated formally and proved in Appendix B.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 $\ell_2$ sphere, i.e., $\mathcal{X} = \mathbb{S}^{d-1}$. Consider $n$ users where user $i$ has some data $x_i \in \mathcal{X}$. For every $i \in [n]$, let $x_i$ be privatized using an $\varepsilon$-LDP mechanism $q(\cdot|x_i)$ and potentially post-processed to obtain an estimate $\hat{x}_i$ of $x_i$. We are interested in estimating the empirical mean $\mu \triangleq \frac{1}{n} \sum x_i$ using $\hat{x}_1, \ldots, \hat{x}_n$ such that the estimation error defined below is minimized

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

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

Bhowmick et al. (2018) show that PrivUnit$_2$ achieves the order-optimal privacy-accuracy trade-off for mean estimation i.e., $r_{\text{ME}}(\hat{\mu}_p, q_{\text{pu}}) = \Theta\left(\frac{d}{\min(\varepsilon, (\varepsilon^2 - 1)^{1/2})}\right)$ where $\hat{\mu}_p \triangleq \frac{1}{n} \sum x_i^{\text{pu}}$. Moreover, compared to other (order-optimal) $\varepsilon$-LDP mean estimation mechanisms, PrivUnit$_2$ admits the best constants and gives the smallest $\ell_2$ error in practice (see Feldman and Talwar (2021)). However, 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 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 $\mu$ and hence cannot get the correct (optimal) order of estimation error in (10). Fortunately, we show (below) that the bias can be corrected by using an estimator which is slightly different compared to the original estimator of PrivUnit$_2$. Further, we also show (below) that the resulting unbiased estimator can simulate PrivUnit$_2$ closely while only using on the order of $\varepsilon$ bits of communication.

### 4.1 Simulating PrivUnit$_2$ using MMRC

Consider the PrivUnit$_2$ $\varepsilon$-LDP mechanism $q_{\text{pu}}$ described in Section 2 with parameters $p_0$ and $\gamma$. PrivUnit$_2$ is a cap-based mechanism with $\text{Cap}_p = \{z \in \mathbb{S}^{d-1} \mid \langle z, x \rangle \geq \gamma\}$ (see Appendix C for details). Let $\pi_{\text{mmrc}}$ be the distribution and $z_1, z_2, \ldots, z_N$ be the candidates obtained from Algorithm 2 when the reference distribution is Unif($\mathbb{S}^{d-1}$). Let $K \sim \pi_{\text{mmrc}}(\cdot)$. 


Define \( p_{\text{mmrc}} \) := \( \mathbb{P}(z_K \in \text{Cap}_x) \) to be the probability with which the sampled candidate \( z_K \) belongs to the spherical cap associated with PrivUnit_2. Define \( m_{\text{mmrc}} \) as the scaling factor in (2) when \( p_0 \) in (2) is replaced by \( p_{\text{mmrc}} \). Define \( \hat{x}_{\text{mmrc}} \) := \( z_K/m_{\text{mmrc}} \) as the estimator of the MMRC mechanism simulating PrivUnit_2. The following Lemma shows that \( \hat{x}_{\text{mmrc}} \) is an unbiased estimator. Proof can be found in Appendix E.1.

**Lemma 4.1.** Let \( \hat{x}_{\text{mmrc}} \) be the estimator of the MMRC mechanism simulating PrivUnit_2 as defined above. Then, \( \mathbb{E}_{q_{\text{mmrc}}}[\hat{x}_{\text{mmrc}}] = x \).

The following Theorem shows that \( q_{\text{mmrc}} \) can compress \( q_{\text{pu}} \) to the order of \( \varepsilon \)-bits of communication as well as simulate it in a near-lossless fashion. Proof can be found in Appendix E.2. The key idea in the proof is to show that when the number of candidates \( N \) is exponential in \( \varepsilon \), the scaling factor \( m_{\text{mmrc}} \) is close to the scaling parameter associated with PrivUnit_2 (i.e., \( m_{\text{pu}} \) defined in (2)).

**Theorem 4.1.** Let \( q_{\text{pu}}(z|x) \) be the \( \varepsilon \)-LDP PrivUnit_2 mechanism with parameters \( p_0 \) and \( \gamma \) and estimator \( \hat{x}_{\text{pu}} \). Let \( q_{\text{mmrc}}(z|x) \) denote the MMRC privatization mechanism simulating PrivUnit_2 with \( N \) candidates and estimator \( \hat{x}_{\text{mmrc}} \) as defined above. Consider any \( \lambda > 0 \). Then,

\[
\mathbb{E}_{q_{\text{mmrc}}}[(\|\hat{x}_{\text{mmrc}} - x\|_2^2) \leq (1 + \lambda)^2 \mathbb{E}_{q_{\text{pu}}}[(\|\hat{x}_{\text{pu}} - x\|_2^2) + 2(1 + \lambda)(2 + \lambda)\sqrt{\mathbb{E}_{q_{\text{pu}}}[(\|\hat{x}_{\text{pu}} - x\|_2^2) + (2 + \lambda)^2]}
\]

as long as

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

Finally, we consider estimating the empirical mean \( \mu \) defined earlier using the MMRC scheme simulating PrivUnit_2. To that end, let \( \bar{x}_{\text{mmrc}} \) be the unbiased estimator of \( x_i \) at the \( i \)th user. Let the estimate of \( \mu \) be \( \hat{\mu}_{\text{mmrc}} := \frac{1}{n} \sum_i \hat{x}_{\text{mmrc}}^i \). For all \( i \in [n] \), since \( \hat{x}_{\text{mmrc}}^i \) are independent of each other as well as unbiased, we obtain the following corollary from Theorem 4.1.

**Corollary 4.1.** Let \( r_{\text{ME}}(\hat{\mu}_{\text{pu}}, q_{\text{pu}}) \) and \( r_{\text{ME}}(\hat{\mu}_{\text{mmrc}}, q_{\text{mmrc}}) \) 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,

\[
r_{\text{ME}}(\hat{\mu}_{\text{mmrc}}, q_{\text{mmrc}}) \leq (1 + \lambda)^2 \mathbb{E}_{q_{\text{pu}}}[(\hat{\mu}_{\text{pu}} - \mu)^2] + 2(1 + \lambda)(2 + \lambda)\sqrt{\mathbb{E}_{q_{\text{pu}}}[(\|\hat{x}_{\text{pu}} - x\|_2^2) + (2 + \lambda)^2]}.
\]

as long as \( N \) satisfies (11).

**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 \( \varepsilon \) 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., \( \mu \neq 0 \). More specifically, we set \( x_1, \ldots, x_{n/2} \overset{i.i.d.}{\sim} N(1,1)^{\otimes d} \) and \( x_{n/2+1}, \ldots, x_n \overset{i.i.d.}{\sim} N(10,1)^{\otimes d} \). Further, to ensure that each data lies on \( S^{d-1} \), we normalize each \( x_i \) by setting \( x_i \leftarrow x_i/\|x_i\|_2 \). See more details and variations in Appendix E.3.

In Figure 1 (left), 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 (right), we show the privacy-accuracy tradeoffs. We see that MMRC simulating PrivUnit_2 can attain the accuracy of the uncompressed PrivUnit_2 for the range of \( \varepsilon \)'s typically considered by LDP mechanisms while only using max\{\( \varepsilon/\ln 2 \) + 2,8\} bits.
Figure 1: Comparing PrivUnit$_2$, MMRC simulating PrivUnit$_2$ and SQKR for mean estimation with $d = 500$ and $n = 5000$. **Left:** $\ell_2$ error vs #bits for $\varepsilon = 6$. **Right:** $\ell_2$ error vs $\varepsilon$ for #bits $= \max\{\varepsilon / \ln 2 \} + 2, 8$. SQKR uses #bits $= \varepsilon$ for both because it leads to a poor performance if #bits $> \varepsilon$.

## 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 $\mathcal{X} := \{e_1, e_2, \ldots, e_d\}$, 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 $x_i \in \mathcal{X}$. For every $i \in [n]$, let $x_i$ be privatized using an $\varepsilon$-LDP mechanism $q(\cdot|x_i)$ and potentially post-processed to obtain an estimate $\hat{x}_i$ of $x_i$. We are interested in estimating the *empirical distribution* of $x_1, \ldots, x_n$, defined as $\Pi \triangleq \frac{1}{n} \sum_i x_i$ using $\hat{x}_1, \ldots, \hat{x}_n$ such that the estimation error defined below is minimized:

$$ r_{\text{FE}}(\hat{\Pi}, \ell, \varepsilon) \triangleq \max_{x^s \in \mathcal{X}^n} \mathbb{E} \left[ \ell \left( \hat{\Pi}(\hat{x}_1, \ldots, \hat{x}_n), \Pi \right) \right], $$

(12)

where $\ell = \|\cdot\|_1$ or $\|\cdot\|_2^2$, $\hat{\Pi}$ is an estimate of $\Pi$ and the expectation is with respect to $q(\cdot|x_i)$ as well as all (possibly shared) randomness used by $q(\cdot|x_i)$ $\forall i \in [n]$. For simplicity, we only focus on $\ell_2$ error i.e., $\ell = \|\cdot\|_2^2$.

Ye and Barg (2018) show that the *Subset Selection* achieves the order-optimal privacy-accuracy trade-off for frequency estimation i.e., $r_{\text{FE}}(\hat{\Pi}_{ss}, \varepsilon) = \Theta\left(\min\left\{\frac{\varepsilon}{\varepsilon(\varepsilon^2 - 1)^{d/2}}, d\right\}\right)$ (where $\hat{\Pi}_{ss} := \frac{1}{n} \sum_i \hat{x}_i^{ss}$). Like PrivUnit$_2$, compared to other (order-optimal) $\varepsilon$-LDP frequency estimation mechanisms, *Subset Selection* admits the best constants and gives the smallest $\ell_2$ error in practice (see Chen et al. (2020)). However, the communication cost associated with *Subset Selection* is $O\left(\frac{d}{\varepsilon(\varepsilon^2 - 1)^{d/2}}\right)$ bits per user, which can be an issue for small and moderate $\varepsilon$.

Similar to PrivUnit$_2$, one could apply the generic MMRC scheme defined in Section 3 to compress and simulate *Subset Selection*. However, for a fixed number of candidates $N$, it yields a biased estimate of $x$ and hence cannot get the correct (optimal) order of estimation error in (12). Fortunately, similar to PrivUnit$_2$, we show (below) that the bias can be corrected by using an estimator which is slightly different compared to the original estimator of *Subset Selection*. Further, we also show (below) that the resulting unbiased estimator can simulate *Subset Selection* closely while only using on the order of $\varepsilon$-bits communication.

### 5.1 Simulating Subset Selection using MMRC

Consider the *Subset Selection* $\varepsilon$-LDP mechanism $q_{ss}$ described in Section 2 with $s := \lceil \frac{d}{\varepsilon + d} \rceil$. *Subset Selection* is cap-based mechanism as discussed in Section 3 and Appendix F with
Consider any $\lambda > 0$ as defined above. Then, any $\lambda > 0$ and $s$.

Subset Selection Theorem 5.1. Let $\text{PrivUnit}$ Appendix H.2. Similar to the previous section.

Empirical Comparisons. Next, we empirically demonstrate the privacy-accuracy-communication tradeoffs. More specifically, MMRC simulating Subset Selection can attain the accuracy of the uncompressed Subset Selection for the range of $\epsilon$’s typically considered by LDP mechanisms while only using $\max\{(|\epsilon|/\ln 2) + 3, 8\}$ bits.

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 $\epsilon$’s typically considered by LDP mechanisms while only using $\max\{(|\epsilon|/\ln 2) + 3, 8\}$ bits.

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

The following Theorem shows that $q^\text{ss}$ can compress $q^\text{ss}$ to the order of $\epsilon$-bits of communication as well as simulate it in a near-lossless fashion. A proof can be found in Appendix H.2. Similar to PrivUnit, the key idea in the proof is to show that when the number of candidates $N$ is exponential in $\epsilon$, the scaling factor $m^\text{mmrc}$ is close to the scaling parameter associated with $q^\text{ss}$ (i.e., $m^\text{ss}$ defined in (4)).

Theorem 5.1. Let $q^\text{ss}(z|x)$ be the $\epsilon$-LDP Subset Selection mechanism with parameters $d$ and $s = \lfloor d/11+\epsilon \rfloor$ and estimator $\hat{x}^\text{ss}$. Let $q^\text{mmrc}(z|x)$ denote the MMRC privatization mechanism simulating Subset Selection with $N$ candidates and estimator $\hat{x}^\text{mmrc}$ as defined above. Consider any $\lambda > 0$. Then,

$$E_{q^\text{mmrc}} \left[ \| \hat{x}^\text{mmrc} - x \|_2^2 \right] \leq (1 + 4\lambda + 5\lambda^2 + 2\lambda^3) E_{q^\text{ss}} \left[ \| \hat{x}^\text{ss} - x \|_2^2 \right],$$

as long as

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

Finally, we consider estimating the empirical frequency $\Pi$ defined earlier using the MMRC scheme simulating Subset Selection. To that end, let $\hat{x}_i^\text{mmrc}$ be the unbiased estimator of $x_i$ at the $i^{th}$ user. Let the estimate of $\Pi$ be $\Pi^\text{mmrc} := \frac{1}{n} \sum_i \hat{x}_i^\text{mmrc}$. For all $i \in [n]$, since $\hat{x}_i^\text{mmrc}$ are independent of each other as well as unbiased, we obtain the following corollary from Theorem 5.1.

Corollary 5.1. Let $r^\text{ff} \left( \Pi^\text{ss}, q^\text{ss} \right)$ and $r^\text{ff} \left( \Pi^\text{mmrc}, q^\text{mmrc} \right)$ be the empirical mean estimation error for Subset Selection and MMRC simulating Subset Selection with $N$ candidates respectively. Consider any $\lambda > 0$. Then

$$r^\text{ff} \left( \Pi^\text{mmrc}, q^\text{mmrc} \right) \leq \left( 1 + 4\lambda + 5\lambda^2 + 2\lambda^3 \right) r^\text{ff} \left( \Pi^\text{ss}, q^\text{ss} \right),$$

as long as $N$ satisfies (13).
Figure 2: Comparing Subset Selection, MMRC simulating Subset Selection and RHR for frequency estimation with $d = 500$ and $n = 5000$. **Top:** $\ell_2$ error vs #bits for $\varepsilon = 6$. **Bottom:** $\ell_2$ error vs $\varepsilon$ for #bits $= \max\{(\varepsilon/\ln 2) + 3, 8\}$. RHR uses #bits $= \varepsilon$ for both because it leads to a poor performance if #bits $> \varepsilon$.

6 Conclusion and Future Work

We demonstrated how Minimal Random Coding can be used to simulate $\varepsilon$-LDP mechanisms in a manner which is communication efficient while preserving accuracy and differential privacy guarantees. However, we did not discuss the computational cost of these methods, which grows exponentially with $\varepsilon$. An important question for future research is therefore how to increase the computational efficiency of MRC and MMRC.

MRC is only one of several schemes which could be considered for the implementation of $\varepsilon$-LDP mechanisms. Other schemes not yet studied for this purpose include Rejection Sampling (Harsha et al., 2007) or the Poisson Functional Representation (Li and El Gamal, 2018).

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. A question left for future research is how much communication is required to establish and select these sources of randomness.

Acknowledgements

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.
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.

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. Konecny, 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. Nguyễn, 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.
Appendix

Organization. The Appendix is organized as follows. In Appendix A, we focus on MRC and provide the proofs of Theorem 3.1, Theorem 3.2, and Theorem 3.3. In Appendix B, we focus on MMRC and provide the proofs of Theorem 3.4 and Theorem 3.5. Further, we also provide Theorem B.1 where we show that MMRC can simulate any $\varepsilon$-LDP cap-based mechanism in a nearly lossless fashion with about $\varepsilon$ bits of communication. In Appendix C, we provide additional preliminary on PrivUnit and also show that PrivUnit2 is a cap-based mechanism (Definition 3.1). In Appendix D, we show how PrivUnit2 can be simulated using MRC analogous to how we simulated PrivUnit2 using MMRC in Section 4.1. Along with the theoretical guarantees, we also provide some empirical comparisons between MRC simulating PrivUnit2 and PrivUnit2. In Appendix E, we provide the proofs of Lemma 4.1 and Theorem 4.1 as well as some additional empirical comparisons between MMRC simulating PrivUnit2 and PrivUnit2. In Appendix F, we provide additional preliminary on Subset Selection and also show that Subset Selection is a cap-based mechanism (Definition 3.1). In Appendix G, we show how Subset Selection can be simulated using MRC analogous to how we simulated Subset Selection using MMRC in Section 5.1. Along with the theoretical guarantees, we also provide some empirical comparisons between MRC simulating Subset Selection and Subset Selection. In Appendix H, 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 Minimal Random Coding

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

$$\pi_{\text{mrc}}(k) := \frac{q(z_k|x)/p(z_k)}{\sum_{k'} q(z_{k'}|x)/p(z_{k'})}$$

$\pi_{\text{mrc}}$ can be viewed as a function that maps $x$ and $(z_1, ..., z_N)$ to a distribution in $[N]$. However for notational convenience, when the context is clear, we will omit the dependence on $x$ and $(z_1, ..., z_N)$.

Let $q_{\text{mrc}}$ denote the distribution of $z_K$ where $K ~ \pi_{\text{mrc}}(\cdot)$ i.e., with $\delta(\cdot)$ denoting the Dirac delta function:

$$q_{\text{mrc}}(z|x) := \sum_k \pi_{\text{mrc}}(k) \delta(z - z_k).$$

A.1 Utility of MRC

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

**Theorem 3.1 (Utility of MRC).** Consider any input alphabet $X$, output alphabet $Z$, data $x \in X$, and $\varepsilon$-LDP mechanism $q(\cdot|x)$. Consider any reference distribution $p(\cdot)$ such that $|\ln(q(z|x)/p(z))| \leq \varepsilon \forall x \in X, z \in Z$.\(^8\) Let the number of candidates be $N = 2^{(\log \varepsilon + \varepsilon c)\varepsilon}$ for some constant $c \geq 0$. Then, $q_{\text{mrc}}$ is such that

$$\left| E_{q_{\text{mrc}}}([z - x]^2) - E_q([z - x]^2) \right| \leq \frac{2\alpha \sqrt{E_q||z - x||^4}}{1 - \alpha}$$

(5)

\(^8\)Note that this condition holds for many reference distributions $p(\cdot)$. For example, one can simple choose $p(\cdot) = q(\cdot|x^*)$ for some $x^* \in X$. 

16
holds with probability at least \(1 - 2\alpha\), where
\[
\alpha = \sqrt{2 - c\varepsilon + 2 - \varepsilon^2 / \log e + 1}.
\]

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 \(t \geq 0\) be some constant and let \(N' = 2(D_{\text{KL}}(q'(z) \| p(z))) + t\). Let \(q\) be a discrete distribution constructed by drawing \(N'\) samples \(\{z_k\}_{k=1}^{N'}\) from \(p\) and defining
\[
q(z) := \sum_{k=1}^{N'} \frac{q(z_k)}{p(z_k)} \delta(z - z_k).
\]
Furthermore, let \(f\) be a measurable function and \(\|f\|_q = \sqrt{\mathbb{E}_{q'(z)}[f^2(z)]}\) be its 2-norm under \(q'\). Then it holds that
\[
\mathbb{P}
\left[
\left|
\mathbb{E}_q[f(z)] - \mathbb{E}_{q'}[f(z)]
\right|
\geq \frac{2\|f\|_q \alpha'}{1 - \alpha'}
\right]
\leq 2\alpha'
\]
where \(\alpha' = \sqrt{2^{-t/4} + 2\sqrt{\mathbb{P}(\log(q'(z)/p(z)) > D_{\text{KL}}(q'(z) \| p(z)) + t/2)}\}.
\]

We apply Theorem 3.2 (Havasi et al., 2019) to \(q'(z) := q(z|x)\) and \(f(z) := \|z - x\|^2\). We identify \(q(z) = q_{\text{rec}}(z|x)\) and \(N' = N\). To prove Theorem 3.1, it suffices to show that \(\alpha \geq \alpha'\). Note that
\[
D_{\text{KL}}(q(z|x) \| p(z)) = \mathbb{E}_{q(z|x)} \left[ \log \left( \frac{q(z|x)}{p(z)} \right) \right] \leq \varepsilon \log e,
\]
where (a) follows the definition of KL-divergence and (b) follows since \(|\log(q(z|x)/p(z))| \leq \varepsilon \log e\) by the assumption on \(p\). We therefore have
\[
t = (\log e + 4c\varepsilon) - D_{\text{KL}}(q(z|x) \| p(z)) \geq 4c\varepsilon.
\]

It follows that
\[
\mathbb{P}(\log(q(z|x)/p(z)) > D_{\text{KL}}(q(z|x) \| p(z)) + t/2)
\leq \mathbb{P}(\log(q(z|x)/p(z)) > \mathbb{E} \left[ \log(q(z|x)/p(z)) \right] + 2c\varepsilon)
\leq \exp(-2c^2/(\log e)^2) = 2^{-2c^2/\log e}.
\]
where (b) follows from Hoeffding’s inequality since \(|\log(q(z|x)/p(z))| \leq \varepsilon \log e\) by the assumption on \(p\). Therefore,
\[
\alpha' \leq \sqrt{2 - c\varepsilon + 2\sqrt{2 - 2c^2/\log e}} = \sqrt{2 - c\varepsilon + 2 - 4c^2/\log e + 1} = \alpha.
\]

Remark A.1. For most \(\varepsilon\)-LDP mechanisms \(q(\cdot|x)\), the term \(\mathbb{E}_q[\|z - x\|^2]\) in (5) can be well-controlled. For instance, for Subset Selection and PrivUnit2, the output spaces are bounded, and therefore, \(\sqrt{\mathbb{E}_q[\|z - x\|^2]}\) is of the same order as \(\mathbb{E}_q[\|z - x\|^2]\). Therefore, by making \(\alpha\) 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.
A.2 Privacy of MRC

A.2.1 Pure Privacy of MRC

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

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

**Proof.** For any $x, x' \in \mathcal{X}, z \in \mathcal{Z}$, using the definition of an $\varepsilon$-LDP mechanism, we have
\[ q(z|x) \leq \exp(\varepsilon)q(z|x'). \]  
(14)

For any $x, x' \in \mathcal{X}$, $\{z_k\}_{k=1}^N \in \mathcal{Z}^N$ and $k \in [N]$, we have
\[
\frac{\pi^{\text{mrc}}(k)}{p^{\text{mrc}}(k)} = (a) \frac{q(z_k|x)}{q(z_k|x')} \times \frac{\sum_{k'} q(z_k'|x')/p(z_k')}{\sum_{k'} q(z_k'|x)/p(z_k')}
\leq \exp(\varepsilon) \times \frac{\sum_{k'} \exp(\varepsilon)q(z_k'|x)/p(z_k')}{\sum_{k'} q(z_k'|x)/p(z_k')}
= \exp(\varepsilon) \times \frac{\sum_{k'} q(z_k'|x)/p(z_k')}{\sum_{k'} q(z_k'|x)/p(z_k')}
= \exp(2\varepsilon).
\]
where (a) follows from the definition of $\pi^{\text{mrc}}$ and $\pi(\cdot)$ and (b) follows from (14). \hfill \Box

A.2.2 Approximate Privacy of MRC

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

**Theorem 3.3** (Approximate DP guarantee of MRC). Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $x \in \mathcal{X}$, and $\varepsilon$-LDP mechanism $q(\cdot|x)$. Consider any reference distribution $p(\cdot)$ such that $|\ln(q(z|x)/p(z))| \leq \varepsilon \ \forall \ x \in \mathcal{X}, z \in \mathcal{Z}$.\footnote{Note that this condition holds for many reference distributions $p(\cdot)$. For example, one can simple choose $p(\cdot) = q(\cdot|x^*)$ for some $x^* \in \mathcal{X}$.} Let $\varepsilon_0 \geq 0$ be some constant and let the number of candidates $N = \exp(2\varepsilon + 2\varepsilon_0)$. Then, for any $\delta \leq 1$, $\pi^{\text{mrc}}(\cdot)$ obtained from Algorithm 1 is $(\varepsilon + \varepsilon_0, \delta)$-LDP mechanism where
\[ \varepsilon_0 := \ln \frac{1 + a_0}{1 - a_0} \quad \text{and} \quad a_0 := \exp(-a_0)\sqrt{\frac{1}{2} \ln \frac{2}{\delta}}. \]

**Proof.** Fix any $x \in \mathcal{X}$. Let us define the following random variable:
\[ w(z|x) = q(z|x)/p(z). \]  
(15)

Assuming $z \sim p(\cdot)$, the expected value of the random variable $w(z|x)$ is
\[ \mathbb{E}_p[w(z|x)] = \mathbb{E}_p[q(z|x)/p(z)] = \int_{z \in \mathcal{Z}} q(z|x) = 1. \]

Further, the random variable $w(z|x)$ can be bounded as follows:
\[ |w(z|x)| = \frac{|q(z|x)/p(z)|}{\exp(\varepsilon)}. \]  
(a)
where (a) follows from the assumption on \( p(\cdot) \). Therefore, we have

\[
\Pr \left( \left| \frac{1}{N} \sum_{k=1}^{N} w(z_k|x) - 1 \right| \geq a_0 \right) \overset{(a)}{\leq} 2 \exp \left( \frac{-2N a_0^2}{(\exp(\varepsilon) - \exp(-\varepsilon))^2} \right) \leq 2 \exp \left( \frac{-2N a_0^2}{\exp(2\varepsilon)} \right) \overset{(b)}{=} \delta
\]

where (a) follows from Hoeffding’s inequality and (b) follows from the definition of \( a_0 \) and \( N \).

Now, for any \( x, x' \in \mathcal{X} \), \( \{z_k\}_{k=1}^{N} \in \mathcal{Z}^{N} \) and \( k \in [N] \), we have

\[
\pi_{\text{mmrc}}^c(k) = \left( \frac{q(z_k|x)}{q(z_k|x')} \right) \times \frac{\sum_{k'} q(z_{k'}|x')/p(z_{k'})}{\sum_{k'} q(z_{k'}|x)/p(z_{k'})} \overset{(b)}{=} \frac{q(z_k|x)}{q(z_k|x')} \times \frac{\sum_{k'} w(z_{k'}|x')}{\sum_{k'} w(z_{k'}|x)} \overset{(c)}{\leq} \exp(\varepsilon) \times \frac{1}{N} \sum_{k'} w(z_{k'}|x') \]

(17)

where (a) follows from the definition of \( \pi_{\text{mmrc}} \), (b) follows from (15) and (c) follows from (14).

Now, using (16) in (17), we have with probability at least \( 1 - \delta \):

\[
\frac{\pi_{\text{mmrc}}^c(k)}{\pi_{\text{mmrc}}^x(k)} \leq \exp(\varepsilon) \times \frac{1 + a_0}{1 - a_0} \overset{(a)}{=} \exp(\varepsilon + \varepsilon_0)
\]

where (a) follows from the definition of \( \varepsilon_0 \).

\[\square\]

**B Modified Minimal Random Coding**

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

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

Similar to \( \pi_{\text{mmrc}} \), \( \pi_{\text{mmrc}}^c \) can be be viewed as a function that maps \( x \) and \( (z_1, ..., z_N) \) to a distribution in \([N] \). However, to reduce clutter, we will generally omit the dependence on \( x \) and \( (z_1, ..., z_N) \). Further, since \( \pi_{\text{mmrc}}^c \) depends on \( (z_1, ..., z_N) \) only through \( \theta \), we will sometimes show this dependence as \( \pi_{\text{mmrc}}^c(\theta) \).

Finally, let \( q_{\text{mmrc}} \) denote the distribution of \( z_K \) where \( K \sim \pi_{\text{mmrc}}^c \). That is, with \( \delta \) denoting the Dirac delta function:

\[
q_{\text{mmrc}}(z|x) := \sum_{k} \pi_{\text{mmrc}}^c(k) \delta(z - z_k).
\]

**B.1 Privacy of MMRC**

In this section, we prove Theorem 3.4 i.e., we show that \( \pi_{\text{mmrc}} \) is a \( \varepsilon \)-LDP mechanism.
**Theorem 3.4** (DP guarantee of MMRC). Consider any input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$, data $x \in \mathcal{X}$, and $\varepsilon$-LDP cap-based mechanism $q(\cdot|x)$. Let the reference distribution $p(\cdot)$ be the uniform distribution on $\mathcal{Z}$. Consider any number of candidates $N \geq 1$. Then, $\pi^{\text{MRC}}(\cdot)$ obtained from Algorithm 2 is an $\varepsilon$-LDP mechanism.

**Proof.** For any $\varepsilon$-LDP cap-based $q(\cdot|x)$, 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 $\pi^{\text{arc}}$ to $\pi^{\text{MRC}}$ ensures that (8) is true, that is,

\[t_l \leq \pi^{\text{MRC}}(k) \leq t_u \forall k \in [N].\]

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

\[
\frac{\pi^{\text{MRC}}_x(k)}{\pi^{\text{MRC}}_{x'}(k)} \leq t_u \frac{(a)}{t_l} \leq \frac{c_1(\varepsilon,d)}{c_2(\varepsilon,d)} \leq \exp(\varepsilon), \tag{b}
\]

where (a) follows from the definitions of $t_u$ and $t_l$ and (b) follows from (18). \hfill \Box

**B.2 Supporting Lemmas to prove the utility of MMRC**

To prove Theorem 3.5 (Section B.3), we prove that the expected KL divergence between $\pi^{\text{arc}}$ and $\pi^{\text{MRC}}$ can be controlled arbitrarily when the number of candidates is of the right order (Lemma B.2). To prove Lemma B.2, we first show that the KL divergence between $\pi^{\text{arc}}$ and $\pi^{\text{MRC}}$, for a given fraction of candidates inside the $\text{Cap}_x$, can be bounded in terms of $\varepsilon$ (Lemma B.1).

**B.2.1 The KL divergence between $\pi^{\text{arc}}$ and $\pi^{\text{MRC}}$ is small**

**Lemma B.1.** Let $q(z|x)$ be an $\varepsilon$-LDP cap-based mechanism. Let $p(z)$ be the uniform distribution over $\mathcal{Z}$ and let $\{z_k\}_{k=1}^N$ be $N$ candidates drawn from $p(z)$. Let $\theta$ denote the fraction of candidates inside the $\text{Cap}_x$ associated with $q(z|x)$. Let $\pi^{\text{arc}}$ be the distribution over the indices $k \in [N]$ under MRC obtained from Algorithm 1 and $\pi^{\text{MRC}}$ be the distribution over the indices $k \in [N]$ under MMRC obtained from Algorithm 2. Then,

\[
D_{\text{KL}} \left( \pi^{\text{arc}}_{x,\theta}(\cdot) \parallel \pi^{\text{arc}}_{x,\theta}(\cdot) \right) \leq \varepsilon \log e.
\]

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

1. For $\theta = E[\theta]$, we have $\pi^{\text{arc}}_{x,\theta}(\cdot) = \pi^{\text{MRC}}_{x,\theta}(\cdot)$. Therefore,

\[
D_{\text{KL}} \left( \pi^{\text{arc}}_{x,\theta}(\cdot) \parallel \pi^{\text{arc}}_{x,\theta}(\cdot) \right) = D_{\text{KL}} \left( \pi^{\text{arc}}(\cdot) \parallel \pi^{\text{arc}}(\cdot) \right) = 0 \leq \varepsilon \log e. \tag{19}
\]

2. If $\theta < E[\theta]$, then $\pi^{\text{arc}}$ violates the upper threshold $t_u$ so that $\pi^{\text{MRC}}(k) = t_u$ for all $k \in \text{Cap}_x$ and we have

\[
D_{\text{KL}} \left( \pi^{\text{arc}}_{x,\theta}(\cdot) \parallel \pi^{\text{ar}}_{x,\theta}(\cdot) \right) = \sum_{k \in \text{Cap}_x} \pi^{\text{arc}}_x(k) \log \frac{\pi^{\text{arc}}_x(k)}{\pi^{\text{MRC}}_x(k)} \tag{a}
\]

\[
= \sum_{k \in \text{Cap}_x} \frac{1}{N} \cdot \frac{c_1(\varepsilon,d)}{c_2(\varepsilon,d)} \log \frac{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \left( \frac{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \right)
\]

\[
+ \sum_{k \notin \text{Cap}_x} \frac{1}{N} \cdot \frac{c_2(\varepsilon,d)}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \left( \frac{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \right).
\]

3. If $\theta > E[\theta]$, then $\pi^{\text{arc}}$ violates the lower threshold $t_l$ so that $\pi^{\text{MRC}}(k) = t_l$ for all $k \in \text{Cap}_x$ and we have

\[
D_{\text{KL}} \left( \pi^{\text{arc}}_{x,\theta}(\cdot) \parallel \pi^{\text{MRC}}_{x,\theta}(\cdot) \right) = \sum_{k \in \text{Cap}_x} \pi^{\text{arc}}_x(k) \log \frac{\pi^{\text{arc}}_x(k)}{\pi^{\text{MRC}}_x(k)} \tag{b}
\]

\[
= \sum_{k \in \text{Cap}_x} \frac{1}{N} \cdot \frac{c_1(\varepsilon,d)}{c_2(\varepsilon,d)} \log \frac{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \left( \frac{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \right)
\]

\[
+ \sum_{k \notin \text{Cap}_x} \frac{1}{N} \cdot \frac{c_2(\varepsilon,d)}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \left( \frac{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))}{c_2(\varepsilon,d) + \theta(c_1(\varepsilon,d) - c_2(\varepsilon,d))} \right).
\]

20
3. For $\theta > \mathbb{E}[\theta]$, we have

$$D_{KL} \left( \pi_{x,\theta}^{\text{mrc}}(\cdot) \bigg| \bigg| \pi_{x,\theta}^{\text{mrc}}(\cdot) \right) = \sum_{x_i} \pi_{x,\theta}^{\text{mrc}}(k) \log \frac{\pi_{x,\theta}^{\text{mrc}}(k)}{\pi_{x,\theta}^{\text{mrc}}(k)}$$

where (a) follows from the definition of $\pi_{x,\theta}^{\text{mrc}}(k)$ and $\pi_{x,\theta}^{\text{mrc}}(k)$, (b) follows because $\{k : k \in \mathbb{C}_{\theta} \} = \theta N$ and $\{k : k \notin \mathbb{C}_{\theta} \} = (1-\theta)N$, (c) follows because $\log \frac{\pi_{x,\theta}^{\text{mrc}}(k)}{\pi_{x,\theta}^{\text{mrc}}(k)} \leq 0$, (d) follows because $\theta \geq 0$, (e) follows because $\mathbb{E}[\theta] \leq 1$, and (f) follows because $\theta_{1}(\varepsilon, d) / (\theta_{2}(\varepsilon, d) \leq \exp(\varepsilon)$.

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

$$D_{KL} \left( \pi_{x,\theta}^{\text{mrc}}(\cdot) \bigg| \bigg| \pi_{x,\theta}^{\text{mrc}}(\cdot) \right) = \sum_{x_i} \pi_{x,\theta}^{\text{mrc}}(k) \log \frac{\pi_{x,\theta}^{\text{mrc}}(k)}{\pi_{x,\theta}^{\text{mrc}}(k)}$$

where (a) follows from the definition of $\pi_{x,\theta}^{\text{mrc}}(k)$ and $\pi_{x,\theta}^{\text{mrc}}(k)$, (b) follows because $\{k : k \in \mathbb{C}_{\theta} \} = \theta N$ and $\{k : k \notin \mathbb{C}_{\theta} \} = (1-\theta)N$, (c) follows because $\log \frac{\pi_{x,\theta}^{\text{mrc}}(k)}{\pi_{x,\theta}^{\text{mrc}}(k)} \leq 0$, (d) follows because $\theta \geq 0$, (e) follows because $\mathbb{E}[\theta] \leq 1$, and (f) follows because $\theta_{1}(\varepsilon, d) / (\theta_{2}(\varepsilon, d) \leq \exp(\varepsilon)$.
\[
\log \frac{c_2(\varepsilon, d)}{c_2(\varepsilon, d) + \theta \times (c_1(\varepsilon, d) - c_2(\varepsilon, d))} \leq 0, \quad (d) \text{ follows because } \theta \leq 1, \quad (e) \text{ follows because } \mathbb{E}[\theta] \geq 0, \quad \text{and (f) follows because } c_1(\varepsilon, d)/c_2(\varepsilon, d) \leq \exp(\varepsilon).
\]

\[\square\]

### B.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 B.2.** Let \( q(z|x) \) be an \( \varepsilon \)-LDP cap-based mechanism. Let \( p(z) \) be the uniform distribution over \( Z \) and let \( \{z_k\}_{k=1}^N \) be \( N \) candidates drawn from \( p(z) \). Let \( \theta \) denote the fraction of candidates inside the Cap associated with \( q(z|x) \). Let \( \pi_{\text{mrc}} \) be the distribution over the indices \( k \in [N] \) under MRC obtained from Algorithm 1 and \( \pi_{\text{mmrc}} \) be the distribution over the indices \( k \in [N] \) under MMRC obtained from Algorithm 2. Then,

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

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

\[
N = \frac{2(\exp(\varepsilon) - 1)^2}{\rho^2} \ln \frac{2}{\rho}.
\]

**Proof.** Let \( \theta \) denote the fraction of candidates inside the cap, i.e.,

\[
\theta = \frac{1}{N} \sum_{k=1}^N I(z_k \in \text{Cap}_x).
\]

Therefore, we have

\[
\mathbb{E}[\theta] = \mathbb{P}_{z_k \sim \text{Unif}(Z)} (z_k \in \text{Cap}_x) = \mathbb{P}_{z \sim \text{Unif}(Z)} (z \in \text{Cap}_x).
\]

(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

\[
\mathbb{E}_\theta \left[ D_{\text{KL}} \left( \pi_{\text{mrc}}^{x, \theta} (\cdot) \middle\| \pi_{\text{mmrc}}^{x, \theta} (\cdot) \right) \right] = \sum_{\theta} \mathbb{P}(\theta) \times D_{\text{KL}} \left( \pi_{\text{mrc}}^{x, \theta} (\cdot) \middle\| \pi_{\text{mmrc}}^{x, \theta} (\cdot) \right)
\]

\[
= \sum_{\theta:|\theta - \mathbb{E}[\theta]| \leq \hat{\rho}} \mathbb{P}(\theta) D_{\text{KL}} \left( \pi_{\text{mrc}}^{x, \theta} (\cdot) \middle\| \pi_{\text{mmrc}}^{x, \theta} (\cdot) \right) + \sum_{\theta:|\theta - \mathbb{E}[\theta]| > \hat{\rho}} \mathbb{P}(\theta) D_{\text{KL}} \left( \pi_{\text{mrc}}^{x, \theta} (\cdot) \middle\| \pi_{\text{mmrc}}^{x, \theta} (\cdot) \right).
\]

(23)

Now, we will upper bound \( D_{\text{KL}} \left( \pi_{\text{mrc}}^{x, \theta} (\cdot) \middle\| \pi_{\text{mmrc}}^{x, \theta} (\cdot) \right) \) whenever \( \theta \) is such that \( |\theta - \mathbb{E}[\theta]| \leq \hat{\rho} \). As in the proof of Lemma B.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 \( D_{\text{KL}} \left( \pi_{\text{mrc}}^{x, \theta} (\cdot) \middle\| \pi_{\text{mmrc}}^{x, \theta} (\cdot) \right) = 0 \).

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

\[
D_{\text{KL}} \left( \pi_{\text{mrc}}^{x, \theta} (\cdot) \middle\| \pi_{\text{mmrc}}^{x, \theta} (\cdot) \right) \leq \log \frac{c_2(\varepsilon, d) + \mathbb{E}[\theta] \times (c_1(\varepsilon, d) - c_2(\varepsilon, d))}{c_2(\varepsilon, d) + \theta \times (c_1(\varepsilon, d) - c_2(\varepsilon, d))} \leq \log \frac{c_2(\varepsilon, d) + \mathbb{E}[\theta] \times (c_1(\varepsilon, d) - c_2(\varepsilon, d))}{c_2(\varepsilon, d) + (\mathbb{E}[\theta] - t) \times (c_1(\varepsilon, d) - c_2(\varepsilon, d))} \leq \log \frac{c_2(\varepsilon, d) + \mathbb{E}[\theta] \times (c_1(\varepsilon, d) - c_2(\varepsilon, d))}{c_2(\varepsilon, d) + (\mathbb{E}[\theta] - t) \times (c_1(\varepsilon, d) - c_2(\varepsilon, d))}
\]
where (a) follows by letting $\theta = E[\theta] - t$ with $t > 0$, (b) follows by using $\log(1 + x) \leq x \log e$ for $x = \frac{t \times (c_1(\varepsilon, d) - c_2(\varepsilon, d))}{1 + (E[\theta] - t) \times (c_1(\varepsilon, d) - c_2(\varepsilon, d))} > 0$, (c) follows because $E[\theta] - t = \theta \geq 0$, and (d) follows because $t = E[\theta] - \theta \leq \hat{\rho}$.

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

$$D_{KL}(\pi_{x,\theta}^{\text{mrc}}(\cdot) \| \pi_{x,\theta}^{\text{mrc}}(\cdot)) \leq \frac{\theta c_1(\varepsilon, d)}{c_2(\varepsilon, d) + \theta (c_1(\varepsilon, d) - c_2(\varepsilon, d))} \log \left( \frac{\theta c_1(\varepsilon, d)}{E[\theta]c_1(\varepsilon, d) + (\theta - E[\theta]) \times c_2(\varepsilon, d)} \right)$$

(a) $$\leq \log \left( \frac{\theta c_1(\varepsilon, d)}{E[\theta]c_1(\varepsilon, d) + E[\theta]c_1(\varepsilon, d) + \theta c_2(\varepsilon, d)} \right)$$

(b) $$= \log \left( \frac{1 + t c_1(\varepsilon, d) - c_2(\varepsilon, d))}{E[\theta]c_1(\varepsilon, d) + t c_2(\varepsilon, d)} \right)$$

(c) $$\leq \frac{\log e \times t (c_1(\varepsilon, d) - c_2(\varepsilon, d))}{E[\theta]c_1(\varepsilon, d) + t c_2(\varepsilon, d)} \leq \frac{\log e \times \hat{\rho} (c_1(\varepsilon, d) - c_2(\varepsilon, d))}{E[\theta]c_1(\varepsilon, d)} \leq \frac{\log e \times \hat{\rho} (c_1(\varepsilon, d) - c_2(\varepsilon, d))}{E[\theta]c_1(\varepsilon, d)}$$

(25)

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

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

$$D_{KL}(\pi_{x,\theta}^{\text{mrc}}(\cdot) \| \pi_{x,\theta}^{\text{mrc}}(\cdot)) \leq \frac{\hat{\rho} (c_1(\varepsilon, d) - c_2(\varepsilon, d))}{\min \{c_2(\varepsilon, d), E[\theta]c_1(\varepsilon, d)\}}$$

(a) $$\leq \frac{\log e \times \hat{\rho} (c_1(\varepsilon, d) - c_2(\varepsilon, d))}{\min \{c_2(\varepsilon, d), c_1(\varepsilon, d)\} \mathbb{P}_{z \sim \text{Unif}(Z)}(z \in \text{Cap}_x)}$$

(b) $$\leq \frac{2 \log e \times \hat{\rho} (c_1(\varepsilon, d) - c_2(\varepsilon, d))}{c_2(\varepsilon, d)}$$

(c) $$\leq 2 \log e \times \hat{\rho} (\exp(\varepsilon) - 1)$$

(26)

where (a) follows from (22), (b) follows because $\mathbb{P}_{z \sim \text{Unif}(Z)}(z \in \text{Cap}_x) \geq c_2(\varepsilon, d)/2c_1(\varepsilon, d)$ 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 B.1 in (23), we have

$$\mathbb{E}_{\theta} \left[ D_{KL}(\pi_{x,\theta}^{\text{mrc}}(\cdot) \| \pi_{x,\theta}^{\text{mrc}}(\cdot)) \right] \leq \sum_{\theta:|\theta-E[\theta]| \leq \hat{\rho}} \mathbb{P}(\theta) \times 2 \log e \times \hat{\rho} (\exp(\varepsilon) - 1) + \sum_{\theta:|\theta-E[\theta]| > \hat{\rho}} \mathbb{P}(\theta) \times \varepsilon \log e$$
\[ (a) \quad 2 \log e \times \hat{\rho} (\exp(\varepsilon) - 1) + \rho \varepsilon \log e \]
\[ (b) \quad 2 \log e \times \sqrt{\frac{\ln(2/\rho)}{2N}} (\exp(\varepsilon) - 1) + \rho \varepsilon \log e \]
\[ (c) \quad \log e \times \rho (1 + \varepsilon) \]
where \( (a) \) follows because \( \mathbb{P} (|\theta - E[\theta]| \leq \hat{\rho}) \leq 1 \) and \( \mathbb{P} (|\theta - E[\theta]| \geq \hat{\rho}) \leq \rho \), \( (b) \) follows by plugging in \( \hat{\rho} = \sqrt{\frac{\ln(2/\rho)}{2N}} \), and \( (c) \) follows by plugging in \( N \).

\section*{B.3 Utility of MMRC}

In this section, we first prove Theorem 3.5 i.e., we show that, with number of candidates exponential in \( \varepsilon \), samples drawn from \( q^\text{mmrc} \) will be similar to the samples drawn from \( q^\text{mrc} \) in terms of \( \ell_2 \) error.

Then, in Theorem B.1, we show that MMRC can simulate any \( \varepsilon \)-LDP cap-based mechanism in a nearly lossless fashion with about \( \varepsilon \) bits of communication.

\subsection*{B.3.1 Utility of MMRC with respect to \( q^\text{mrc} \)}

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

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

where \( \rho \in (0,1) \) is such that

\[ N = \frac{2(\exp(\varepsilon) - 1)^2}{\rho^2} \ln \frac{2}{\rho}. \]

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

\[ \mathbb{E}_{q^\text{mmrc}}[||z - x||^2_2] - \mathbb{E}_{q^\text{mrc}}[||z - x||^2_2] \leq \max_{x,z} ||z - x||^2_2 \times ||q^\text{mrc}(z|x) - q^\text{mmrc}(z|x)||_{\text{TV}}. \]

Next, we will upper bound the total variation distance between \( q^\text{mrc} \) and \( q^\text{mmrc} \) using Pinsker’s inequality as follows:

\[ ||q^\text{mrc}(z|x) - q^\text{mmrc}(z|x)||_{\text{TV}} \leq \sqrt{\frac{1}{2 \log e} D_{\text{KL}} (q^\text{mrc}(z|x) || q^\text{mmrc}(z|x)).} \]

Next, we will upper bound the KL divergence between \( q^\text{mrc}(z|x) \) and \( q^\text{mmrc}(z|x) \). To that end, for every \( x \in \mathcal{X} \), let \( p^\text{mrc}(z_1, \ldots, z_N, K, z_K|x) \) denote the joint distribution of the candidates \( z_1, \ldots, z_N \) drawn from \( p(z) \), the transmitted index \( K \) under MRC, and the sample \( z_K \) corresponding to \( K \). We have

\[ p^\text{mrc}(z_1, \ldots, z_N, K, z_K|x) = p(z_1, \ldots, z_N|x) \times p^\text{mrc}(K|z_1, \ldots, z_N, x) \times p^\text{mrc}(z_K|z_1, \ldots, z_N, K, x) \]
\[ = p(z_1, \ldots, z_N|x) \times p^\text{mrc}(K|z_1, \ldots, z_N, x) \times p^\text{mrc}(z_K|z_1, \ldots, z_N, K, x) \]
\[ = p(z_1, \ldots, z_N|x) \times p^\text{mrc}(z_K|z_1, \ldots, z_N, K, x) \]
where (a) follows because \( z_1, \ldots, z_N \) are independent of \( x \), (b) follows because \( p_{\text{mrc}}(K | z_1, \ldots, z_N, x) = \pi_{x, \theta}(k) \), and (c) follows because \( p_{\text{mrc}}(z_K | z_1, \ldots, z_N, K, x) = 1 \) (note that \( z_K \) can be viewed as a function of \( (z_1, \ldots, z_N, K) \)).

Similarly, for every \( x \in X \), let \( q_{\text{mrc}}(z_1, \ldots, z_N, K, z_K | x) \) denote the joint distribution of the candidates \( z_1, \ldots, z_N \) drawn from \( p(z) \), the transmitted index \( K \) under MMRC, and the sample \( z_K \) corresponding to \( K \). We have

\[
q_{\text{mrc}}(z_1, \ldots, z_N, K, z_K | x) = p(z_1, \ldots, z_N | x) \times p_{\text{mrc}}(K | z_1, \ldots, z_N, x) \times p_{\text{mrc}}(z_K | z_1, \ldots, z_N, K, x)
\]

(a) follows because \( z_1, \ldots, z_N \) are independent of \( x \), (b) follows because \( p_{\text{mrc}}(K | z_1, \ldots, z_N, x) = \pi_{x, \theta}(k) \), and (c) follows because \( p_{\text{mrc}}(z_K | z_1, \ldots, z_N, K, x) = 1 \).

We are now in a position to upper bound the KL divergence between \( q_{\text{mrc}}(z_K | x) \) and \( q_{\text{mrc}}(z_K | x) \):

\[
D_{\text{KL}}(q_{\text{mrc}}(z | x) \parallel q_{\text{mrc}}(z | x)) \leq D_{\text{KL}}(p_{\text{mrc}}(z_1, \ldots, z_N, K, z_K | x) \parallel \pi_{x, \theta}(k))
\]

(a) follows because the chain rule for KL-divergence, (b) follows from (29) and (30), (c) follows by the definition of KL-divergence, (d) follows because \( \pi_{x, \theta} \) and \( \pi_{x, \theta} \) depend on \( z_1, \ldots, z_N \) only via \( \theta \) for cap-based mechanisms, and (e) follows from Lemma B.2 because \( N = \frac{2(\exp(1) - 1)^2}{\rho^2} \ln \frac{2}{\rho} \). Combining (27), (28), and (31), we have

\[
\mathbb{E}_{q_{\text{mrc}}} [\|z - x\|_2^2] \leq \mathbb{E}_{q_{\text{mrc}}} [\|z - x\|_2^2] + \sqrt{\frac{\rho(1 + \varepsilon)}{2} \times \max_{x,z} \|z - x\|_2^2} + \frac{2\alpha}{1 - \alpha} \times \sqrt{\mathbb{E}_{q} [\|z - x\|_2^4]}
\]

Remark B.1. For bounded \( \varepsilon \)-LDP mechanisms such as PrivUnit2 and Subset Selection, the term \( \max_{x,z} \|z - x\|_2^2 \) in (9) is of the same order as \( \mathbb{E}_{q} [\|z - x\|_2^2] \). Therefore, by picking a large \( N \) in Theorem 3.5 (i.e. \( \log N \geq C\varepsilon \) for a sufficiently large \( C \)), \( \rho \) can be made arbitrarily small and the estimation error of MMRC can be arbitrarily close to the estimation error of MRC.

B.3.2 Utility of MMRC with respect to \( q \)

Theorem B.1. Consider any input alphabet \( X \), output alphabet \( Z \), data \( x \in X \), and \( \varepsilon \)-LDP cap-based mechanism \( q(\cdot | x) \). Let the reference distribution \( p(\cdot) \) be the uniform distribution on \( Z \). Let \( N \) denote the number of candidates. Then, \( q_{\text{mrc}} \) is such that

\[
\mathbb{E}_{q_{\text{mrc}}} [\|z - x\|_2^2] \leq \mathbb{E}_{q} [\|z - x\|_2^2] + \sqrt{\frac{\rho(1 + \varepsilon)}{2} \times \max_{x,z} \|z - x\|_2^2} + \frac{2\alpha}{1 - \alpha} \times \sqrt{\mathbb{E}_{q} [\|z - x\|_2^4]}
\]
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

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

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

C Preliminary on PrivUnit$_2$

First, we briefly recap the PrivUnit$_2$ mechanism ($q_{pu}^a$) proposed in Bhowmick et al. (2018). PrivUnit$_2$ is a private sampling scheme when the input alphabet $\mathcal{X}$ is the $d$-dimensional unit $\ell_2$ sphere $S^{d-1}$. More formally, given a vector $x \in S^{d-1}$, PrivUnit$_2$ (see Algorithm 3) draws a vector $z$ from a spherical cap $\{z \in S^{d-1} | \langle z, x \rangle \geq \gamma \}$ with probability $p_0 \geq 1/2$ or from its complement $\{z \in S^{d-1} | \langle z, x \rangle < \gamma \}$ 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 $q_{pu}^a(z | x)$ is:

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

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

**Algorithm 3: Privatized Unit Vector: PrivUnit$_2$**

**Require:** $x \in S^{d-1}$, $\gamma \in [0, 1]$, $p_0 \geq 1/2$.

Draw random vector $z$ according to the distribution

$$z = \begin{cases} 
    \text{uniform on } \{z \in S^{d-1} | \langle z, x \rangle \geq \gamma \} & \text{with probability } p_0 \\
    \text{uniform on } \{z \in S^{d-1} | \langle z, x \rangle < \gamma \} & \text{otherwise.}
\end{cases}$$

Set $\alpha = \frac{d - 1}{2}$, $\tau = \frac{1 + \gamma}{2}$, and

$$m_{pu} = \frac{(1 - \gamma^2)^\alpha}{2^{d-2}(d - 1)} \left[ \frac{p_0}{B(\alpha, \alpha) - B(\tau; \alpha, \alpha)} - \frac{1 - p_0}{B(\tau; \alpha, \alpha)} \right] \quad (33)$$

**return** $\hat{x}_{pu} = \frac{z}{m_{pu}}$

Given its inputs $x, \gamma$, and $p_0$, Algorithm 3 returns an estimator $\hat{x}_{pu} := z/m_{pu}$ which is differentially private and unbiased where $m_{pu}$ is a scaling factor. The choice of $\gamma$ described in Theorem C.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.
The following theorem borrowed from Bhowmick et al. (2018) describes the choice of $\gamma$ and provides the precise associated differential privacy guarantee of the PrivUnit$_2$ mechanism.

**Theorem C.1 (Bhowmick et al. (2018)).** Let $\gamma \in [0,1]$ and $p_0 = \frac{e^\varepsilon_0}{1+e^\varepsilon_0}$. Then algorithm PrivUnit$_2(\cdot, \gamma, p_0)$ is $\varepsilon = (\varepsilon + \varepsilon_0)$-differentially private whenever $\gamma \geq 0$ is such that

$$
\varepsilon \geq \log \frac{1 + \gamma \cdot \sqrt{2(d-1)/\pi}}{(1 - \gamma \cdot \sqrt{2(d-1)/\pi})}, \text{ i.e. } \gamma \leq \frac{e^\varepsilon - 1}{e^\varepsilon + 1} \sqrt{\frac{\pi}{2(d-1)}},
$$

or

$$
\varepsilon \geq \frac{1}{2} \log(d) + \log(6) - \frac{d}{2} \log(1 - \gamma^2) + \log \gamma \quad \text{and} \quad \gamma \geq \sqrt{\frac{2}{d}}.
$$

(34)

Here, $\varepsilon$ can be viewed as the total privacy budget. Typically, $\mu$ fraction of this budget is allocated for the spherical cap threshold $\gamma$ and $1 - \mu$ fraction is allocated to the probability parameter $p_0$ with which a particular spherical cap is chosen i.e., $\varepsilon = \mu \varepsilon$ and $\varepsilon_0 = (1 - \mu) \varepsilon$ for some $\mu \in [0,1]$. While the parameter $\mu$ 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 PrivUnit$_2$ can be easily extended to the setup where $\mu$ needs to be optimized over.

**C.2 PrivUnit$_2$ is unbiased and order-optimal**

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

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

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

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

**C.3 PrivUnit$_2$ is a cap-based mechanism**

The randomness in the estimator $\hat{x}^{Pu}$ obtained from the PrivUnit$_2(x, \gamma, p_0)$ mechanism comes from $z$. Therefore, we obtain a convenient expression for the conditional distribution of $z$ conditioned on $x$ i.e., $q^{Pu}(z|x)$. Define $\text{Cap}_x := \{z|(x, z) \geq \gamma\}$. Recall from (34) that $\gamma$ is a function of $\varepsilon$ and $d$. Further, as described in Section C.1, when the budget split parameter $\mu$ is known, $p_0$ can viewed as a function of $\varepsilon$. Then, the conditional distribution $q^{Pu}(z|x)$ in (32) can be written as follows:

$$
q^{Pu}(z|x) = \begin{cases} 
  c_1(\varepsilon, d) & \text{if } z \in \text{Cap}_x \\
  c_2(\varepsilon, d) & \text{if } z \notin \text{Cap}_x
\end{cases}
$$

(35)

where $c_1(\varepsilon, d) = p_0 \times \frac{2}{A(1,d)I_1 - \gamma^2(d+1,\frac{1}{2})}$ and $c_2(\varepsilon, d) = (1-p_0) \times \frac{2}{A(1,d) - A(1,d)I_1 - \gamma^2(d+1,\frac{1}{2})}$ are functions of $\varepsilon$ and $d$. 

27
Further, $P_{z \sim \text{Unif}(Z)}(z \in \mathcal{Z}_x) = \frac{I_1 - e^{-d/2}}{2}$. Therefore,

$$\frac{c_1(\varepsilon, \delta)}{c_2(\varepsilon, \delta)} \times P_{z \sim \text{Unif}(Z)}(z \in \mathcal{Z}_x) = \frac{p_0 \times (2 - I_1 - e^{-d/2})}{2(1 - p_0)} \overset{(a)}{\geq} \frac{p_0}{2(1 - p_0)} \overset{(b)}{=} \frac{1}{2},$$

where (a) follows because $I_1 - e^{-d/2} \leq 1$ and (b) follows because $p_0 \geq 1/2$.

## D Simulating PrivUnit\textsubscript{2} using Minimal Random Coding

In this section, we simulate PrivUnit\textsubscript{2} using MRC analogous to how we simulate PrivUnit\textsubscript{2} using MMRC in Section 4.1. First, in Appendix D.1, we provide an unbiased estimator for MRC simulating PrivUnit\textsubscript{2}. Next, in Appendix D.2 we provide the utility guarantee associated with MRC simulating PrivUnit\textsubscript{2}. To do that, first, in Appendix D.2.1, we show that when the number of candidates $N$ is exponential in $\varepsilon$, the scaling factor $m_{\text{arc}}$ is close to the scaling parameter associated with PrivUnit\textsubscript{2} (i.e., $m_{\text{pu}}$). Next, in Appendix D.2.2, we show that if the scaling factor $m_{\text{arc}}$ is close to the scaling parameter $m_{\text{pu}}$, then the mean squared error associated with MRC simulating PrivUnit\textsubscript{2} is close to the mean squared error associated with PrivUnit\textsubscript{2}. In Appendix D.2.3, we combine everything and show that $q_{\text{arc}}$ can compress $q_{\text{pu}}$ to the order of $\varepsilon$-bits of communication as well as simulate it in a near-lossless fashion. Finally, in Appendix D.3, we provide some empirical comparisons.

### D.1 Unbiased Minimal Random Coding simulating PrivUnit\textsubscript{2}

Consider the PrivUnit\textsubscript{2} $\varepsilon$-LDP mechanism $q_{\text{pu}}$ described in Section 2 with parameters $p_0$ and $\gamma$. PrivUnit\textsubscript{2} is a cap-based mechanism with $\text{Cap}_x = \{z \in S^{d-1} \mid \langle z, x \rangle \geq \gamma\}$ as discussed in Appendix C. Let $\pi_{\text{arc}}$ be the distribution and $z_1, z_2, \ldots, z_N$ be the candidates obtained from Algorithm 1 when the reference distribution is Unif($S^{d-1}$). Let $K \sim \pi_{\text{arc}}(\cdot)$. Define $p_{\text{arc}} := P(Z_K \in \text{Cap}_x)$ to be the probability with which the sampled candidate $z_K$ belongs to the spherical cap associated with PrivUnit\textsubscript{2}. Define $m_{\text{arc}}$ as the scaling factor in (2) when $p_0$ in (2) is replaced by $p_{\text{arc}}$. Define $\hat{x}_{\text{arc}} := z_K / m_{\text{arc}}$ as the estimator of the MRC mechanism simulating PrivUnit\textsubscript{2}. The following Lemma shows that $\hat{x}_{\text{arc}}$ is an unbiased estimator.

**Lemma D.1.** Let $\hat{x}_{\text{arc}}$ be the estimator of the MRC mechanism simulating PrivUnit\textsubscript{2} as defined above. Then, $E_{q_{\text{arc}}}[\hat{x}_{\text{arc}}] = x$.

**Proof.** For $k \in [N]$, let $A_k := 1(z_k \in \text{Cap}_x)$. Then, $p_{\text{arc}} = P(A_K = 1)$. Using the definition of $\hat{x}_{\text{arc}}$, we have

$$E_{q_{\text{arc}}}[\hat{x}_{\text{arc}}] = 1/m_{\text{arc}} E_{q_{\text{arc}}}[z_K].$$

Let us evaluate $E_{q_{\text{arc}}}[z_K]$. We have

$$E_{q_{\text{arc}}}[z_K] \overset{(a)}{=} E_{K, z_1, \ldots, z_N}[z_K] \overset{(b)}{=} E_{z_1, \ldots, z_N} \sum_{k=1}^{N} \pi_{\text{arc}}[z_1, \ldots, z_N] \langle z_k \rangle \times z_k \overset{(c)}{=} E_{A_1, \ldots, A_N} \left[ E_{z_1, \ldots, z_N} \left[ \sum_{k=1}^{N} \pi_{\text{arc}}[z_1, \ldots, z_N] \langle z_k \rangle \times z_k \mid A_1, \ldots, A_N \right] \right] \overset{(d)}{=} \sum_{k=1}^{N} E_{A_1, \ldots, A_N} \left[ E_{z_1, \ldots, z_N} \left[ \pi_{\text{arc}}[z_1, \ldots, z_N] \langle z_k \rangle \times z_k \mid A_1, \ldots, A_N \right] \right].$$
where (a) follows because the randomness in $q_{\text{arc}}$ comes from the randomness in $K, z_1, \ldots, z_N$, (b) follows by calculating the expectation over $K$ and showing the dependence of $\pi_{\text{arc}}$ on $z_1, \ldots, z_N$, explicitly, (c) follows by the tower property of expectation, (d) follows by linearity of expectation, (e) follows because $\pi_{\text{arc}, z_1, \ldots, z_N}(k) = \pi_{\text{arc}, A_1, \ldots, A_N}(k)$ since $\pi_{\text{arc}}$ depends on $z_1, \ldots, z_N$ via $A_1, \ldots, A_N$, (f) follows because $z_k$ is independent of $A_1, \ldots, A_{k-1}, A_{k+1}, \ldots, A_N$ given $A_k$, (g) follows by marginalizing $z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_N$, (h) follows by the tower property of expectation, (i) follows by evaluating the expectation over $A_k$, (j) follows because $E_x[z | A = 1] := E_z[z | A = 1]$ and $E_x[z | A = 0] := E_z[z | A = 0]$ are constants for every $k \in [N]$, (k) follows by marginalizing $A_1, \ldots, A_N$, (l) follows from the definitions of $P(A_K = 1)$ and $P(A_K = 0)$, and (m) follows from rotational symmetry (see the proof of Lemma 4.1 in Bhowmick et al. (2018) for details). Therefore, we can write

$$E_{q_{\text{arc}}}[\hat{x}_{\text{arc}}] = \frac{1}{m_{\text{arc}}} E_{q_{\text{arc}}}[z_K] \overset{(a)}{=} x$$

where (a) follows from (36).
D.2 Utility of Minimal Random Coding simulating PrivUnit$_2$

D.2.1 The scaling factors of PrivUnit$_2$ 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 $\varepsilon$, then the scaling parameters associated with PrivUnit$_2$ and the MRC scheme simulating PrivUnit$_2$ are close.

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

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

as long as

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

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

$$\gamma_+ \triangleq \frac{(1 - \gamma^2)^\alpha}{2^{d-2}(d-1) (B(\alpha, \alpha) - B(\tau; \alpha, \alpha))}, \quad \text{and} \quad \gamma_- \triangleq \frac{(1 - \gamma^2)^\alpha}{2^{d-2}(d-1) (B(\tau; \alpha, \alpha))}.$$ Therefore, we have

$$\frac{1}{m_{\text{mrc}}} - \frac{1}{m_{\text{pu}}} = \frac{m_{\text{pu}} - m_{\text{mrc}}}{m_{\text{pu}} \cdot m_{\text{mrc}}} = \frac{1}{m_{\text{pu}}} \frac{(\gamma_+ - \gamma_-) \cdot (p_0 - p_{\text{mrc}})}{((\gamma_+ - \gamma_-) p_{\text{mrc}} + \gamma_-)}$$

$$= \frac{1}{m_{\text{pu}}} \left( \frac{p_0 - p_{\text{mrc}}}{p_{\text{mrc}} + \gamma_-} \right)$$

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

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

(38)

We will now upper bound $\frac{p_0 - p_{\text{mrc}}}{p_0 - 1/2}$. We start by obtaining convenient expressions for $p_{\text{mrc}}$ and $p_0$. To compute $p_{\text{mrc}} = \mathbb{P}(z_K \in \text{Cap}_x)$, recall that $\theta$ denotes the fraction of candidates that belong inside the $\text{Cap}_x$. 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 $\mathbb{P}(z_K \in \text{Cap}_x) = \bar{c}_1(\varepsilon, d)/p_0$. Further, since $z_k$ are generated uniformly at random,

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

so we have

$$p_{\text{mrc}} = \mathbb{P} \{ z_K \in \text{Cap}_x \} = \mathbb{E} [ \mathbb{P} \{ z_K \in \text{Cap}_x | \theta \} ]$$

30
where \((a)\) follows from (7) because \(q^\text{pa}\) is a cap-based mechanism and \((b)\) follows by simple manipulations.

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

\[
\frac{p_0}{\tilde{c}_1(\varepsilon, d)} + \frac{1 - p_0}{\tilde{c}_2(\varepsilon, d)} = 1
\]

Using this and with some simple manipulations, we have

\[
p_0 = \frac{\tilde{c}_1(\varepsilon, d)\tilde{c}_2(\varepsilon, d)}{(\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d))^2} \left( \frac{\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d)}{\tilde{c}_2(\varepsilon, d)} - \frac{1}{\mathbb{E}[\theta] + \frac{1}{\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d)}} \right).
\]

From (39) and (40), we have

\[
p_0 - p_{\text{hr}} = \frac{\tilde{c}_1(\varepsilon, d)\tilde{c}_2(\varepsilon, d)}{(\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d))^2} \left( \mathbb{E} \left[ \frac{1}{\theta + \frac{\tilde{c}_2(\varepsilon, d)}{\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d)}} - \frac{1}{\mathbb{E}[\theta] + \frac{\tilde{c}_2(\varepsilon, d)}{\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d)}} \right] \right)
\]

\[
= \frac{\tilde{c}_1(\varepsilon, d)\tilde{c}_2(\varepsilon, d)}{(\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d))^2} \left( \mathbb{E} \left[ \theta + \frac{\tilde{c}_2(\varepsilon, d)}{\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d)} \right] - \mathbb{E}[\theta] \right) \left( \mathbb{E}[\theta] + \frac{\tilde{c}_2(\varepsilon, d)}{\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d)} \right)
\]

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

\[
p_0 - p_{\text{hr}} \leq \frac{\tilde{c}_1(\varepsilon, d)\tilde{c}_2(\varepsilon, d)}{(\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d))^2} \times
\]

\[
\left( \frac{p_0}{\tilde{c}_1(\varepsilon, d)} + \frac{\tilde{c}_2(\varepsilon, d)}{\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d)} \right) + \beta
\]

where we have also plugged in \(\mathbb{E}[\theta] = \frac{p_0}{\tilde{c}_1(\varepsilon, d)}\). Now, we can lower bound

\[
\left( \frac{p_0}{\tilde{c}_1(\varepsilon, d)} + \frac{\tilde{c}_2(\varepsilon, d)}{\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d)} \right) \geq \frac{\tilde{c}_2(\varepsilon, d)}{\tilde{c}_1(\varepsilon, d) - \tilde{c}_2(\varepsilon, d)} \geq \frac{1}{\exp(\varepsilon) - 1}
\]
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

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

(42)

Using (42) in (41), we have

$$p_0 - p_{\text{arc}} \leq \frac{\bar{c}_1(\varepsilon,d)\bar{c}_2(\varepsilon,d)}{(\bar{c}_1(\varepsilon,d) - \bar{c}_2(\varepsilon,d))^2} \times (\frac{p_0}{\bar{c}_1(\varepsilon,d)} + \frac{\bar{c}_2(\varepsilon,d)}{\bar{c}_1(\varepsilon,d) - \bar{c}_2(\varepsilon,d)}) \leq \beta$$

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

$$\beta \leq \frac{\lambda(p_0 - 1/2)}{2(1 + \lambda)},$$

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

(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 $\frac{p_0 - p_{\text{arc}}}{p_0 - 1/2}$ i.e.,

$$\frac{p_0 - p_{\text{arc}}}{p_0 - 1/2} \leq \frac{\lambda}{1 + \lambda}.$$

(45)

Using (45) in (38), we have

$$\frac{1}{m_{\text{arc}}} - \frac{1}{m_{\text{pu}}} \leq \frac{\lambda}{m_{\text{pu}}}.$$

(46)

Rearranging (46) completes the proof.

\[\square\]

**D.2.2 Relationship between the scaling factors and mean squared errors associated with PrivUnit\textsubscript{2} and MRC simulating PrivUnit\textsubscript{2}**

In the following Proposition, we show that if the scaling factor $m_{\text{arc}}$ is close to the scaling parameter $m_{\text{pu}}$, then the mean squared error associated with MRC simulating PrivUnit\textsubscript{2} is close to the mean squared error associated with PrivUnit\textsubscript{2}.
Proposition D.1. Let \( q^{\text{pu}}(z|\mathbf{x}) \) be the \( \varepsilon \)-LDP PrivUnit\( _2 \) mechanism with parameters \( p_0 \) and \( \gamma \) and estimator \( \hat{\mathbf{x}}^{\text{pu}} \). Let \( q^{\text{mrc}}(z|\mathbf{x}) \) denote the MRC privatization mechanism simulating PrivUnit\( _2 \) with \( N \) candidates and estimator \( \hat{\mathbf{x}}^{\text{mrc}} \). Let \( m_{\text{pu}} \) denote the scaling factor associated with PrivUnit\( _2 \) and \( m_{\text{mrc}} \) denote the scaling factor associated with the MRC scheme simulating PrivUnit\( _2 \). Consider any \( \lambda > 0 \). If \( m_{\text{pu}} - m_{\text{mrc}} \leq \lambda \cdot m_{\text{mrc}} \), then

\[
\mathbb{E}_{q^{\text{mrc}}} \left[ \| \hat{\mathbf{x}}^{\text{mrc}} - \mathbf{x} \|^2 \right] \leq (1 + \lambda)^2 \mathbb{E}_{q^{\text{pu}}} \left[ \| \hat{\mathbf{x}}^{\text{pu}} - \mathbf{x} \|^2 \right] + 2(1 + \lambda)(2 + \lambda) \sqrt{\mathbb{E}_{q^{\text{pu}}} \left[ \| \hat{\mathbf{x}}^{\text{pu}} - \mathbf{x} \|^2 \right]} + (2 + \lambda)^2.
\]

Proof. We will start by upper bounding \( 1/m_{\text{pu}} \) in terms of \( \mathbb{E}_{q^{\text{pu}}} \left[ \| \hat{\mathbf{x}}^{\text{pu}} - \mathbf{x} \|^2 \right] \). First, observe that

\[
\| \hat{\mathbf{x}}^{\text{pu}} - \mathbf{x} \| \geq \| \mathbf{x} \| \geq \frac{1}{m_{\text{pu}}} - 1 \tag{47}
\]

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

\[
\frac{1}{m_{\text{pu}}} = \frac{1}{m_{\text{pu}}} - 1 + (a) \leq \sqrt{\mathbb{E}_{q^{\text{pu}}} \left[ \| \hat{\mathbf{x}}^{\text{pu}} - \mathbf{x} \|^2 \right]} + 1 \tag{48}
\]

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

\[
\mathbb{E}_{q^{\text{mrc}}} \left[ \| \hat{\mathbf{x}}^{\text{mrc}} - \mathbf{x} \|^2 \right] = \mathbb{E}_{q^{\text{mrc}}} \left[ \| \hat{\mathbf{x}}^{\text{mrc}} \|^2 \right] + \| \mathbf{x} \|^2 + 2 \mathbb{E}_{q^{\text{mrc}}} \left[ \| \hat{\mathbf{x}}^{\text{mrc}} \| \cdot \| \mathbf{x} \| \right]
\]

\[
\leq (a) \mathbb{E}_{q^{\text{mrc}}} \left[ \| \hat{\mathbf{x}}^{\text{mrc}} \|^2 \right] + \| \mathbf{x} \|^2 + 2 \mathbb{E}_{q^{\text{mrc}}} \left[ \| \hat{\mathbf{x}}^{\text{mrc}} \| \cdot \| \mathbf{x} \| \right]
\]

\[
\leq (b) \left( \frac{1}{m_{\text{mrc}}} \right)^2 + 1 + \frac{2}{m_{\text{mrc}}}
\]

\[
\leq (c) \left( \frac{1 + \lambda}{m_{\text{pu}}} \right)^2 + 1 + \frac{2(1 + \lambda)}{m_{\text{pu}}}
\]

\[
\leq (d) (1 + \lambda)^2 \mathbb{E}_{q^{\text{pu}}} \left[ \| \hat{\mathbf{x}}^{\text{pu}} - \mathbf{x} \|^2 \right] + 2(1 + \lambda)(2 + \lambda) \sqrt{\mathbb{E}_{q^{\text{pu}}} \left[ \| \hat{\mathbf{x}}^{\text{pu}} - \mathbf{x} \|^2 \right]} + (2 + \lambda)^2
\]

where (a) follows from Cauchy–Schwarz inequality, (b) follows because \( \| \hat{\mathbf{x}}^{\text{mrc}} \| = 1/m_{\text{mrc}} \) and \( \| \mathbf{x} \| \leq 1 \), (c) follows because \( m_{\text{pu}} - m_{\text{mrc}} \leq \lambda \cdot m_{\text{mrc}} \), and (d) follows using (48) and some simple manipulations. \( \square \)

D.2.3 Simulating PrivUnit\( _2 \) using Minimal Random Coding

The following Theorem shows that \( q^{\text{mrc}} \) can compress \( q^{\text{pu}} \) to the order of \( \varepsilon \)-bits of communication as well as simulate it in a near-lossless fashion.

Theorem D.1. Let \( q^{\text{pu}}(z|\mathbf{x}) \) be the \( \varepsilon \)-LDP PrivUnit\( _2 \) mechanism with parameters \( p_0 \) and \( \gamma \) and estimator \( \hat{\mathbf{x}}^{\text{pu}} \). Let \( q^{\text{mrc}}(z|\mathbf{x}) \) denote the MRC privatization mechanism simulating PrivUnit\( _2 \) with \( N \) candidates and estimator \( \hat{\mathbf{x}}^{\text{mrc}} \). Consider any \( \lambda > 0 \). Then,

\[
\mathbb{E}_{q^{\text{mrc}}} \left[ \| \hat{\mathbf{x}}^{\text{mrc}} - \mathbf{x} \|^2 \right] \leq (1 + \lambda)^2 \mathbb{E}_{q^{\text{pu}}} \left[ \| \hat{\mathbf{x}}^{\text{pu}} - \mathbf{x} \|^2 \right] + 2(1 + \lambda)(2 + \lambda) \sqrt{\mathbb{E}_{q^{\text{pu}}} \left[ \| \hat{\mathbf{x}}^{\text{pu}} - \mathbf{x} \|^2 \right]} + (2 + \lambda)^2
\]

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 D.1 and Lemma D.2. \( \square \)
D.3 Empirical Comparisons

In this section, we compare MRC simulating PrivUnit2 (using its approximate DP guarantee) against PrivUnit2 and SQKR for mean estimation with $d = 500$ and $n = 5000$. We use the same data generation scheme described in Section 4.1 and set $\delta = 10^{-6}$. As before, SQKR uses $\#\text{-bits} = \varepsilon$ because it leads to a poor performance if $\#\text{-bits} > \varepsilon$. We show the privacy-accuracy tradeoffs for these three methods in Figure 3. We see that MRC simulating PrivUnit2 can attain the accuracy of the uncompressed PrivUnit2 for the range of $\varepsilon$’s typically considered by LDP mechanisms while only using $(3\varepsilon/\ln 2) + 6$ bits. In comparison with the results from Section 4.1, the results in this section come with an approximate guarantee ($\delta = 10^{-6}$) and with a higher number of bits of communication.

E Modified Minimal Random Coding Simulating PrivUnit2

In this section, we prove Lemma 4.1 and Theorem 4.1. To prove Theorem 4.1, first, in Appendix E.2.1, we show that when the number of candidates $N$ is exponential in $\varepsilon$, the scaling factor $m_{\text{mmrc}}$ is close to the scaling parameter associated with PrivUnit2 (i.e., $m_{\text{pu}}$). Next, in Appendix E.2.2, we show that if the scaling factor $m_{\text{mmrc}}$ is close to the scaling parameter $m_{\text{pu}}$, then the mean squared error associated with MMRC simulating PrivUnit2 is close to the mean squared error associated with PrivUnit2. Finally, in Appendix E.3, we provide some empirical comparisons in addition to the ones in Section 4.1 between MMRC simulating PrivUnit2 and PrivUnit2.

E.1 Unbiased Modified Minimal Random Coding simulating PrivUnit2

Consider the PrivUnit2 $\varepsilon$-LDP mechanism $q^{\text{pu}}$ described in Section 2 with parameters $p_0$ and $\gamma$. PrivUnit2 is a cap-based mechanism with $\text{Cap}_x = \{z \in S^{d-1} | \langle z, x \rangle \geq \gamma \}$ as discussed in Appendix C. Let $\pi_{\text{mmrc}}$ be the distribution and $z_1, z_2, ..., z_N$ be the candidates obtained from Algorithm 2 when the reference distribution is $\text{Unif}(S^{d-1})$. Let $K \sim \pi_{\text{mmrc}}(\cdot)$. Define $p_{\text{mmrc}} := \mathbb{P}(z_K \in \text{Cap}_x)$ to be the probability with which the sampled candidate $z_K$ belongs to the spherical cap associated with PrivUnit2. Define $m_{\text{mmrc}}$ as the scaling factor in (2) when $p_0$ in (2) is replaced by $p_{\text{mmrc}}$. Define $\hat{x}_{\text{mmrc}} := z_K/m_{\text{mmrc}}$ as the estimator of the MMRC mechanism simulating PrivUnit2.

Lemma 4.1. Let $\hat{x}_{\text{mmrc}}$ be the estimator of the MMRC mechanism simulating PrivUnit2 as defined above. Then, $\mathbb{E}_{q^{\text{mmrc}}} [\hat{x}_{\text{mmrc}}] = x$.

Proof. The proof is similar to the proof of Lemma D.1. \qed
E.2 Utility of Modified Minimal Random Coding simulating PrivUnit₂

E.2.1 The scaling factors of PrivUnit₂ 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 \( \varepsilon \), then the scaling parameters associated with PrivUnit₂ and the MMRC scheme simulating PrivUnit₂ are close.

**Lemma E.1.** Let \( N \) denote the number of candidates used in the MMRC scheme. Let \( K \sim \pi^\text{mmrc} \) where \( \pi^\text{mmrc} \) is the distribution over the indices \([N]\) associated the MMRC scheme simulating PrivUnit₂(\( x, \gamma, p₀ \)). Consider any \( \lambda > 0 \). Then, the scaling factor \( m_{pu} \) associated with PrivUnit₂ and the scaling factor \( m^\text{mmrc} \) associated with the MMRC scheme simulating PrivUnit₂ are such that

\[
m_{pu} - m^\text{mmrc} \leq \lambda \cdot m^\text{mmrc}
\]

as long as

\[
N \geq \frac{e^{2\varepsilon}}{2} \left( \frac{2(1 + \lambda)}{\lambda (p₀ - 1/2)} \right)^2 \ln \left( \frac{4(1 + \lambda)}{\lambda (p₀ - 1/2)} \right).
\]

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

\[
\frac{1}{m^\text{mmrc}} - \frac{1}{m_{pu}} \leq \frac{1}{m_{pu}} \left( \frac{1}{p₀ - 1/2} \right) \left( \frac{1}{p₀ - m^\text{mmrc}} - 1 \right)
\]

We will now upper bound \( \frac{p₀ - m^\text{mmrc}}{p₀ - 1/2} \). We start by obtaining expressions for \( p^\text{mmrc} \) and \( p₀ \).

To compute \( p^\text{mmrc} := \mathbb{P}\{z_K \in \text{Cap}_x\} \), recall that \( \theta \) denotes the fraction of candidates that belong inside the \( \text{Cap}_x \). Let \( c₁(\varepsilon, d) \) and \( c₂(\varepsilon, d) \) be as defined in (35). Let \( \bar{c}_1(\varepsilon, d) = c₁(\varepsilon, d) \times A(1, d) \) and \( \bar{c}_2(\varepsilon, d) = c₂(\varepsilon, d) \times A(1, d) \). It is easy to see from Algorithm 3 and (35) that \( \mathbb{P}(z_k \in \text{Cap}_x) = \bar{c}_1(\varepsilon, d)/p₀ \). Further, since \( z_k \) are generated uniformly at random,

\[
\theta \sim \frac{1}{N} \text{Binom} \left( N, \frac{\bar{c}_1(\varepsilon, d)}{p₀} \right),
\]

so we have

\[
p^\text{mmrc} = \mathbb{P}\{z_K \in \text{Cap}_x\} = \mathbb{E}[\mathbb{P}\{z_K \in \text{Cap}_x|\theta\}]
= \mathbb{E} \left[ \frac{\theta \bar{c}_1(\varepsilon, d)}{\bar{c}_2(\varepsilon, d) + \mathbb{E}[\theta] \bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)} \times 1(\theta \leq \mathbb{E}[\theta]) + \mathbb{E}[\theta] \bar{c}_1(\varepsilon, d) + (\theta - \mathbb{E}[\theta]) \bar{c}_2(\varepsilon, d) \times 1(\theta > \mathbb{E}[\theta]) \right]
\tag{49}
\]

where \((a)\) follows from Algorithm 2.

Similarly, with some simple manipulations on the definition of \( p₀ \), we have

\[
p₀ = \frac{\mathbb{E}[\theta] \bar{c}_1(\varepsilon, d)}{\bar{c}_2(\varepsilon, d) + \mathbb{E}[\theta] \bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)}
\tag{50}
\]

From (49) and (50), we have

\[
p₀ - p^\text{mmrc} = \frac{\mathbb{E}[\bar{c}_1(\varepsilon, d)](\mathbb{E}[\theta] - \theta) \times 1(\theta \leq \mathbb{E}[\theta]) + \bar{c}_2(\varepsilon, d)(\mathbb{E}[\theta] - \theta) \times 1(\theta > \mathbb{E}[\theta])}{\bar{c}_2(\varepsilon, d) + \mathbb{E}[\theta] \bar{c}_1(\varepsilon, d) - \bar{c}_2(\varepsilon, d)}
\]
Let \( \hat{\theta} \) and estimator \( \tilde{\theta} \) with \( \hat{\theta} = \tilde{\theta} \).

\[
\begin{align*}
\rho(\hat{\theta}, \tilde{\theta}) &\leq \frac{\mathbb{E}[\hat{c}(\epsilon, d)](\mathbb{E}[\epsilon] - \theta) \times 1(\theta \leq \mathbb{E}[\epsilon])}{\mathbb{E}[\hat{c}(\epsilon, d)] + \mathbb{E}[\epsilon](\hat{c}(\epsilon, d) - \hat{c}(\epsilon, d))} \\
&\leq \frac{\mathbb{E}[\hat{c}(\epsilon, d)](\mathbb{E}[\epsilon] - \theta) \times 1(\theta > \mathbb{E}[\epsilon])}{\mathbb{E}[\hat{c}(\epsilon, d)] + \mathbb{E}[\epsilon](\hat{c}(\epsilon, d) - \hat{c}(\epsilon, d))}.
\end{align*}
\]

where \( a \) follows because \( (\mathbb{E}[\epsilon] - \theta) \times 1(\theta > \mathbb{E}[\epsilon]) \leq 0 \). Now, using the Hoeffding’s inequality, we have \( \mathbb{P}\{\|\hat{\theta} - \mathbb{E}[\theta]\| < \frac{\ln(2/\beta)}{2N}\} \leq \beta \). Conditioned on the event \( \{\|\hat{\theta} - \mathbb{E}[\theta]\| \leq \frac{\ln(2/\beta)}{2N}\} \) and using the fact that \( |p_{0} - p_{\text{mmrc}}| \leq 1 \), we have

\[
p_{0} - p_{\text{mmrc}} \leq \frac{\hat{c}(\epsilon, d)\sqrt{\ln(2/\beta)}/2N + \beta}{\hat{c}(\epsilon, d)\sqrt{\ln(2/\beta)/2N}} + \beta \leq \lambda(p_{0} - 1/2) \frac{1}{1 + \lambda}.
\]

where \( a \) follows because \( \mathbb{E}[\epsilon] \geq 0 \), \( b \) follows because \( \hat{c}(\epsilon, d)/\hat{c}(\epsilon, d) \leq e^{\epsilon} \), and \( c \) follows if we pick

\[
\beta \leq \frac{\lambda(p_{0} - 1/2)}{2(1 + \lambda)},
\]

\[
N \geq \frac{\exp(2\epsilon)\ln(2/\beta)}{\lambda(p_{0} - 1/2) - \beta} = \frac{\exp(2\epsilon)}{2} \left( \frac{2(1 + \lambda)}{\lambda(p_{0} - 1/2)} \right) \ln \left( \frac{4(1 + \lambda)}{\lambda(p_{0} - 1/2)} \right).
\]

The rest of the proof is similar to the proof of Lemma D.2.

E.2.2 Relationship between the scaling factors and mean squared errors associated with PrivUnit2 and MMRC simulating PrivUnit2

In the following Proposition, we show that if the scaling factor \( m_{\text{mmrc}} \) is close to the scaling parameter \( m_{\text{pu}} \), then the mean squared error associated with MMRC simulating PrivUnit2 is close to the mean squared error associated with PrivUnit2.

**Proposition E.1.** Let \( q_{\text{pu}}(z|x) \) be the \( \epsilon \)-LDP PrivUnit2 mechanism with parameters \( p_{0} \) and \( \gamma \) and estimator \( \hat{\epsilon}^{\text{pu}} \). Let \( q_{\text{mmrc}}(z|x) \) denote the MMRC privatization mechanism simulating PrivUnit2 with \( N \) candidates and estimator \( \hat{\epsilon}^{\text{mmrc}} \). Let \( m_{\text{pu}} \) denote the scaling factor associated with PrivUnit2 and \( m_{\text{mmrc}} \) denote the scaling factor associated with the MMRC scheme simulating PrivUnit2. Consider any \( \lambda > 0 \). If \( m_{\text{pu}} - m_{\text{mmrc}} \leq \lambda \cdot m_{\text{mmrc}} \), then

\[
\mathbb{E}_{q_{\text{mmrc}}} \left[ \|\hat{\epsilon}^{\text{mmrc}} - \bar{x}\|^{2} \right] \leq (1 + \lambda)^{2} \mathbb{E}_{q_{\text{pu}}} \left[ \|\hat{\epsilon}^{\text{pu}} - \bar{x}\|^{2} \right] + 2(1 + \lambda)(2 + \lambda)\sqrt{\mathbb{E}_{q_{\text{pu}}} \left[ \|\hat{\epsilon}^{\text{pu}} - \bar{x}\|^{2} \right]} + (2 + \lambda)^{2}.
\]

**Proof.** The proof is similar to the proof of Proposition D.1.

E.2.3 Simulating PrivUnit2 using Modified Minimal Random Coding

Now, we provide a proof of Theorem 4.1.

**Theorem 4.1.** Let \( q_{\text{pu}}(z|x) \) be the \( \epsilon \)-LDP PrivUnit2 mechanism with parameters \( p_{0} \) and \( \gamma \) and estimator \( \hat{\epsilon}^{\text{pu}} \). Let \( q_{\text{mmrc}}(z|x) \) denote the MMRC privatization mechanism simulating PrivUnit2 with \( N \) candidates and estimator \( \hat{\epsilon}^{\text{mmrc}} \) as defined above. Consider any \( \lambda > 0 \). Then,

\[
\mathbb{E}_{q_{\text{mmrc}}} \left[ \|\hat{\epsilon}^{\text{mmrc}} - \bar{x}\|^{2} \right] \leq (1 + \lambda)^{2} \mathbb{E}_{q_{\text{pu}}} \left[ \|\hat{\epsilon}^{\text{pu}} - \bar{x}\|^{2} \right] + 2(1 + \lambda)(2 + \lambda)\sqrt{\mathbb{E}_{q_{\text{pu}}} \left[ \|\hat{\epsilon}^{\text{pu}} - \bar{x}\|^{2} \right]} + (2 + \lambda)^{2}
\]

as long as

\[
N \geq \frac{2\epsilon}{\lambda} \left( \frac{2(1 + \lambda)}{\lambda(p_{0} - 1/2)} \right)^{2} \ln \left( \frac{4(1 + \lambda)}{\lambda(p_{0} - 1/2)} \right). \tag{11}
\]

**Proof.** The proof follows from Proposition E.1 and Lemma E.1.
E.3 Additional Empirical Comparisons

In Section 4.1, we empirically demonstrated the privacy-accuracy-communication tradeoffs of MMRC simulating PrivUnit2 against PrivUnit2 and SQKR in terms of $\ell_2$ error vs #bits and $\ell_2$ error vs $\varepsilon$ (see Figure 1). In this section, we provide comparisons between these methods in terms of $\ell_2$ error vs $d$ (see Figure 4 (left)) and $\ell_2$ error vs $n$ (see Figure 4 (right)) for a fixed $\varepsilon (=6)$ 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 PrivUnit2, MMRC simulating PrivUnit2 and SQKR for mean estimation with $\varepsilon = 6$ and #bits = 11. Left: $\ell_2$ error vs $d$ for $n = 5000$. Right: $\ell_2$ error vs $n$ for $d = 500.
which is order-optimal. Moreover, if we pick $s$

where

The final estimator of $x$ is denoted by $\hat{x}^{ss}$ and is defined as $\frac{1}{m_{ss}} \cdot (z - b_{ss})$. In other words, $m_{ss}$ and $b_{ss}$ are used to bias the outcome $z$. The scheme is summarized in Algorithm 4.

**Algorithm 4: Subset Selection**

Require: $x \in [d], s \in [d]$.

Draw a $s$-hot random vector $z$ according to the distribution $q^{ss}(z|x)$ in (52).

return $\hat{x}^{ss} = \frac{1}{m_{ss}} \cdot (z - b_{ss})$

---

**F.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 F.1.** Let $\hat{x}^{ss} = \text{Subset Selection}(x, s)$ for some $x \in \mathbb{S}^{d-1}$ and $s \in [d]$. Then, $\mathbb{E}[\hat{x}^{ss}] = x$. Further, the $l_2$ estimation error is

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

Moreover, if we pick $s := \lceil \frac{d}{1 + \varepsilon^2} \rceil$, then

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

which is order-optimal.

**F.2 Subset Selection is a cap-based mechanism**

As discussed in Section 3, $q^{ss}$ defined in (3) is a cap-based mechanism with $\text{Cap}_x = Z_x$, $c_1(\varepsilon, d) = \frac{\varepsilon^2}{(d-1)\varepsilon^2 + (d-1)s}$, and $c_2(\varepsilon, d) = \frac{s}{(d-1)\varepsilon^2 + (d-1)s}$.

Further, $\mathbb{P}_{z \sim \text{Unif}(Z)}(z \in Z_x) = \frac{(d-1)\varepsilon^2}{s} = \frac{s}{d}$. Therefore,

$$
c_1(\varepsilon, d) \times \mathbb{P}_{z \sim \text{Unif}(Z)}(z \in Z_x) = \varepsilon^2 \times s \left( \frac{a}{d} \right) = \frac{\varepsilon^2}{d} \times \left[ \frac{d}{1 + \varepsilon^2} \right] \geq \frac{\varepsilon^2}{d} \times \frac{d}{1 + \varepsilon^2} \geq \frac{1}{2}
$$

where (a) follows by plugging in $s = \lceil \frac{d}{1 + \varepsilon^2} \rceil$ and (b) follows because $\varepsilon \geq 0$. 

38
G 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 5.1. First, in Appendix G.1, we provide an unbiased estimator for MRC simulating Subset Selection. Next, in Appendix G.2 we provide the utility guarantee associated with MRC simulating Subset Selection. To do that, first, in Appendix G.2.1, we show that when the number of candidates \( N \) is exponential in \( \varepsilon \), the scaling factor \( m_{\text{mrc}} \) is close to the scaling parameter associated with Subset Selection (i.e., \( m_{\text{ss}} \)). Next, in Appendix G.2.2, we show that if \( m_{\text{mrc}} \) is close to \( m_{\text{ss}} \) and \( b_{\text{mrc}} \geq b_{\text{ss}} \), then the mean squared error associated with MRC simulating Subset Selection is close to the mean squared error associated with Subset Selection. In Appendix G.2.3, we combine everything and show that \( q_{\text{mrc}} \) can compress \( q_{\text{ss}} \) to the order of \( \varepsilon \)-bits of communication as well as simulate it in a near-lossless fashion. Finally, in Appendix G.3, we provide some empirical comparisons.

G.1 Unbiased Minimal Random Coding simulating Subset Selection

Consider the Subset Selection \( \varepsilon \)-LDP mechanism \( q_{\text{ss}} \) with parameter \( s \) as described in Section 2 and Appendix F. Let \( \pi_{\text{mrc}} \) be the distribution and \( z_1, z_2, \ldots, z_N \) be the candidates obtained from Algorithm 1 when the reference distribution is \( \text{Unif}(Z) \) where \( Z \) is defined in (51). Let \( \theta \) denote the fraction of candidates inside \( \text{Cap}_x = \tilde{Z}_x \) where \( \tilde{Z}_x \) is the set of elements in \( Z \) with 1 in the same location as \( x \). It is easy to see that \( \theta \sim \frac{1}{N} \text{Binom} \left( N, \frac{1}{s} \right) \). Let \( q_{\text{mrc}} = P(z_i = 1) \) where \( z \sim q_{\text{mrc}}(\cdot|x) \) i.e., \( q_{\text{mrc}} = \mathbb{P}\{(z_K)_i = 1\} \) where \( K \sim \pi_{\text{mrc}}(\cdot) \).

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

**Lemma G.1.** Let \( K \sim \pi_{\text{mrc}}(\cdot) \) and \( q_{\text{mrc}} = \mathbb{P}\{(z_K)_i = 1\} \) for \( i \in [d] \). Then,

\[
q_{\text{mrc}} = p_im_{\text{mrc}} + b_{\text{mrc}}
\]

where

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

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

Further, \( \tilde{x}_{\text{mrc}} := (z_K - b_{\text{mrc}})/m_{\text{mrc}} \) is an unbiased estimator of \( x \), i.e., \( \mathbb{E}[\tilde{x}_{\text{mrc}}] = x \).

**Proof.** We have

\[
\mathbb{P}\{(z_K)_i = 1\} = \sum_j p_j \mathbb{P}\{(z_K)_i = 1|x = j\} (a) = \frac{1}{p_i} \mathbb{P}\{(z_K)_i = 1|x = i\} + (1 - p_i) \mathbb{P}\{(z_K)_i = 1|x = j\}.
\]

(56)

where \((a)\) follows by symmetry. Next, we compute \( \mathbb{P}\{(z_K)_i = 1|x = i\} \) and \( \mathbb{P}\{(z_K)_i = 1|x = j\} \) separately.

To compute \( \mathbb{P}\{(z_K)_i = 1|x = i\} \), recall that \( \theta \) denotes the fraction of candidates that belong inside the \( \text{Cap}_x \) i.e., have 1 in the same location as \( x \). From Appendix F.2, recall that \( c_1(\varepsilon, d) := \frac{e^\varepsilon}{(d - 1)e^\varepsilon + (d - 1)s} \), \( c_2(\varepsilon, d) := \frac{1}{(d - 1)e^\varepsilon + (d - 1)s} \). Further, since \( z_k \) are generated uniformly at random,

\[
\theta \sim \frac{1}{N} \text{Binom} \left( N, \frac{(d - 1)s}{d} \right) = \frac{1}{N} \text{Binom} \left( N, \frac{s}{d} \right).
\]
so we have
\[
\mathbb{P}\{(z_K)_i = 1 | x = i\} = \mathbb{P}\{z_K \in \text{Cap}_x | x = i\} = \mathbb{E}[\mathbb{P}\{z_K \in \text{Cap}_x | x = i, \theta\}] = \mathbb{E}\left[\frac{c_1(\varepsilon, d)\theta}{c_1(\varepsilon, d)\theta + (1 - \theta)c_2(\varepsilon, d)}\right]
\]
\[
= \mathbb{E}\left[\frac{e^{\varepsilon \theta}}{e^{\varepsilon \theta} + (1 - \theta)}\right],
\]
(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 $\mathbb{P}\{(z_K)_i = 1 | x = j\}$, we decompose it into
\[
\mathbb{P}\{(z_K)_i = 1 | x = j\} = \mathbb{P}\{(z_K)_i = 1, (z_K)_j = 1 | x = j\} + \mathbb{P}\{(z_K)_i = 1, (z_K)_j = 0 | x = j\}
\]
for any $j \neq i$ and calculate each of the terms separately.

As before, let $\theta$ denotes the fraction of candidates that belong inside the $\text{Cap}_x$ i.e., have 1 in the same location as $x$. Further, let $\bar{\theta}$ denotes the fraction of candidates that belong inside the $\text{Cap}_x$ i.e., have 1 in the same location as $x$ as well as have 1 in the $j^{th}$ location. Since $z_k$ are generated uniformly at random,
\[
\bar{\theta} \sim \frac{1}{N}\text{Binom}\left(N\theta, \frac{(d-2)}{(d-1)}\right) = \frac{1}{N}\text{Binom}\left(N\bar{\theta}, \frac{s-1}{d-1}\right),
\]
so we have
\[
\mathbb{P}\{(z_K)_i = 1, (z_K)_j = 1 | x = j\} = \mathbb{E}_\theta[\mathbb{E}_{\bar{\theta}}[\mathbb{P}\{(z_K)_i = 1, (z_K)_j = 1 | x = j, \bar{\theta}, \theta\}]]
\]
\[
= \mathbb{E}_\theta\left[\mathbb{E}_{\bar{\theta}}\left[\frac{c_1(\varepsilon, d)\bar{\theta}}{c_1(\varepsilon, d)\theta + (1 - \theta)c_2(\varepsilon, d)}\right]\right]
\]
\[
= \mathbb{E}_\theta\left[\frac{s - 1}{d - 1}\mathbb{E}_{\bar{\theta}}\left[\frac{c_1(\varepsilon, d)\bar{\theta}}{c_1(\varepsilon, d)\theta + (1 - \theta)c_2(\varepsilon, d)}\right]\right]
\]
\[
= \mathbb{E}_\theta\left[\frac{s - 1}{d - 1}\mathbb{E}\left[\frac{e^{\varepsilon \theta}}{e^{\varepsilon \theta} + (1 - \theta)}\right]\right].
\]
(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 $\mathbb{P}\{(z_K)_i = 1, (z_K)_j = 0 | x = j\}$, let $\bar{\theta}$ denote the fraction of candidates that belong inside the $\text{Cap}_x$ i.e., have 1 in the same location as $x$ as well as have 0 in the $j^{th}$ location. Since $z_k$ are generated uniformly at random,
\[
\bar{\theta} \sim \frac{1}{N}\text{Binom}\left(N(1 - \theta), \frac{(d-2)}{(d-1)}\right) = \frac{1}{N}\text{Binom}\left(N(1 - \theta), \frac{s}{d-1}\right),
\]
so we have
\[
\mathbb{P}\{(z_K)_i = 1, (z_K)_j = 0 | x = j\} = \mathbb{E}_\theta[\mathbb{E}_{\bar{\theta}}[\mathbb{P}\{(z_K)_i = 1, (z_K)_j = 0 | x = j, \bar{\theta}, \theta\}]]
\]
\[
= \mathbb{E}_\theta\left[\mathbb{E}_{\bar{\theta}}\left[\frac{c_2(\varepsilon, d)\bar{\theta}}{c_1(\varepsilon, d)\theta + (1 - \theta)c_2(\varepsilon, d)}\right]\right]
\]
\[
= \mathbb{E}_\theta\left[\frac{s}{d - 1}\mathbb{E}_{\bar{\theta}}\left[\frac{c_2(\varepsilon, d)(1 - \theta)}{c_1(\varepsilon, d)\theta + (1 - \theta)c_2(\varepsilon, d)}\right]\right]
\]
\[
= \mathbb{E}_\theta\left[\frac{s}{d - 1}\mathbb{E}\left[\frac{(1 - \theta)}{e^{\varepsilon \theta} + (1 - \theta)}\right]\right].
\]
(60)
where (a) follows by the law of total probability, (b) follows because $\mathbb{E}[\theta] = \frac{s}{d-1} \times (1 - \theta)$, and (c) is due to $c_1(\varepsilon, d)/c_2(\varepsilon, d) = e^\varepsilon$. Using (59) and (60) in (58), we have

$$
\mathbb{P}\{ (z_K)_i = 1 | x = j \} = \mathbb{P}\{ (z_K)_i = 1, (z_K)_j = 1 | x = j \} + \mathbb{P}\{ (z_K)_i = 1, (z_K)_j = 0 | x = j \}
$$

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

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

(61)

Combining everything, we have

$$
q_{i}^{\text{src}} = \mathbb{P}\{ (z_K)_i = 1 \}
$$

(a) $p_i \times \left[ \mathbb{P}\{ (z_K)_i = 1 | x = i \} - \mathbb{P}\{ (z_K)_i = 1 | x = j \} \right] + \mathbb{P}\{ (z_K)_i = 1 | x = j \}$.

(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)

(c) $= p_i m_{\text{src}} + b_{\text{src}}$

where (a) follows from (56), (b) follows from (57) and (61), and (c) follows from the definitions of $m_{\text{src}}$ and $b_{\text{src}}$.

Note that the above conclusion holds for all prior distribution $p = (p_1, ..., p_d)$ such that $x \sim p$. Thus by setting $p = x$ (here $x$ is viewed as a one-hot vector), i.e., letting $p$ be the point mass distribution at $x$, we have

$$
\mathbb{E}[x_{\text{src}}] = (\mathbb{E}[z_K] - b_{\text{src}})/m_{\text{src}} = (q^{\text{src}} - b_{\text{src}})/m_{\text{src}} = ((m_{\text{src}} \cdot p + b_{\text{src}}) - b_{\text{src}})/m_{\text{src}} = p \stackrel{(a)}{=} x,
$$

where (a) is due to our construction of $p$. \qed

### G.2 Utility of Minimal Random Coding simulating Subset Selection

#### G.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 $\varepsilon$, then the scaling parameters associated with Subset Selection and the MRC scheme simulating Subset Selection are close.

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

$$
m_{\text{ss}} - m_{\text{src}} \leq \lambda \cdot m_{\text{src}}
$$

and $b_{\text{ss}} \leq b_{\text{src}}$ 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 $m_{\text{ss}}$ and $b_{\text{ss}}$ defined in (55). We can write

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

(62)
\[
b_{ss} := \frac{1}{d - 1} \left( s - \frac{e^\varepsilon \mathbb{E}[\theta]}{e^\varepsilon \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \right). \\
(63)
\]

To verify these, we simply plug \( \mathbb{E}[\theta] = \frac{s}{d} \) into (62) resulting in:
\[
m_{ss} = \frac{d}{d - 1} \frac{se^\varepsilon}{se^\varepsilon + (d - s)} - \frac{s}{d - 1} = \frac{dse^\varepsilon - s^2e^\varepsilon - s(d - t)}{(d - 1)(se^\varepsilon + d - s)} = \frac{s(d - s)(e^\varepsilon - 1)}{(d - 1)(se^\varepsilon + d - s)}.
\]

and into (63) resulting in:
\[
b_{ss} = \frac{1}{d - 1} \left( s - \frac{se^\varepsilon}{se^\varepsilon + d - s} \right) = \frac{1}{d - 1} \left( \frac{s^2e^\varepsilon + s(d - s) - se^\varepsilon}{se^\varepsilon + d - s} \right) = \frac{1}{d - 1} \left( \frac{s(s - 1)e^\varepsilon + s(d - s)}{se^\varepsilon + d - s} \right).
\]

Recall the definitions of \( b_{ss} \) and \( m_{ss} \) from Lemma G.1. Applying Jensen’s inequality on the concave function \( x \mapsto \frac{x}{e^x} \) for some \( c > 0 \) yields \( m_{\text{arc}} \leq m_{ss} \) and \( b_{\text{arc}} \geq b_{ss} \).

Now, we will bound \( |m_{\text{arc}} - m_{ss}| \):
\[
|m_{\text{arc}} - m_{ss}| = \left| \left( \frac{d}{d - 1} \right) \left( \frac{\mathbb{E}[\theta] e^\varepsilon}{e^\varepsilon \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} - \mathbb{E} \left[ \frac{\theta e^\varepsilon}{e^\varepsilon \theta + (1 - \theta)} \right] \right) \right|
\]
\[
\leq 2 \left( \frac{\mathbb{E}[\theta] e^\varepsilon}{e^\varepsilon \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} - \mathbb{E} \left[ \frac{\theta e^\varepsilon}{e^\varepsilon \theta + (1 - \theta)} \right] \right)
\]
\[
= 2 \left( \mathbb{E} \left[ \frac{(\mathbb{E}[\theta] - \theta) e^\varepsilon}{((\mathbb{E}[\theta] - 1) \mathbb{E}[\theta] + 1)((\mathbb{E}[\theta] - 1) \mathbb{E}[\theta] + 1)} \right] \right) \right| \right|
\]
\[
= 2 \left( \mathbb{E} \left[ \frac{(\mathbb{E}[\theta] - \theta) e^\varepsilon}{((\mathbb{E}[\theta] - 1) \mathbb{E}[\theta] + 1)((\mathbb{E}[\theta] - 1) \mathbb{E}[\theta] + 1)} \right] - \frac{(\mathbb{E}[\theta] - \theta) e^\varepsilon}{((\mathbb{E}[\theta] - 1) \mathbb{E}[\theta] + 1)((\mathbb{E}[\theta] - 1) \mathbb{E}[\theta] + 1)} \right| \right|
\]
\[
\leq 2 \frac{\mathbb{E}[\theta] e^\varepsilon}{((\mathbb{E}[\theta] - 1) \mathbb{E}[\theta] + 1)((\mathbb{E}[\theta] - 1) \mathbb{E}[\theta] + 1)} \right| \right|
\]
\[
\leq 4 \mathbb{E} \left[ \frac{(\mathbb{E}[\theta] - \theta) e^\varepsilon}{((\mathbb{E}[\theta] - 1) \mathbb{E}[\theta] + 1)^2} \right] + 2 \beta \leq 4 \sqrt{\frac{\ln(2/\beta)}{2N}} \frac{e^\varepsilon(1 + e^\varepsilon)^2}{4e^{2\varepsilon}} + 2 \beta
\]
\[
= \sqrt{\frac{\ln(2/\beta)}{2N}} \left( e^\varepsilon + 2 + \frac{1}{e^\varepsilon} \right) + 2 \beta \leq \sqrt{\frac{\ln(2/\beta)}{2N}} (e^\varepsilon + 3) + 2 \beta,
\]

where \((a)\) holds since
\[
\frac{(\mathbb{E}[\theta] - \theta) e^\varepsilon}{((\mathbb{E}[\theta] - 1) \mathbb{E}[\theta] + 1)((\mathbb{E}[\theta] - 1) \mathbb{E}[\theta] + 1)} = \frac{\mathbb{E}[\theta] e^\varepsilon}{e^\varepsilon \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} - \frac{\theta e^\varepsilon}{e^\varepsilon \theta + (1 - \theta)} \leq 1,
\]

\((b)\) holds if we pick \(N\) large enough so that \(|\theta - \mathbb{E}[\theta]| \leq \frac{\mathbb{E}[\theta]}{2}\) for which a sufficient condition is \(\sqrt{\frac{\ln(2/\beta)}{2N}} \leq \frac{\mathbb{E}[\theta]}{2}\) i.e., \(N \geq 2 \frac{\ln(2/\beta)}{\mathbb{E}[\theta]^2} = 2(d/s)^2 \ln(2/\beta)\), and \((c)\) holds since \(\mathbb{E}[\theta] = s/d \geq 1/(1 + e^\varepsilon)\). Notice that the constraint \(N \geq 2(d/s)^2 \ln(2/\beta)\) in inequality \((b)\) can be further satisfied as long as \(N \geq 2 \ln(2/\beta)(1 + e^\varepsilon)^2\) since \(s/d \geq 1/(1 + e^\varepsilon)\).
Next, we lower bound \( m_{ss} \) in (62):

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

where (a) holds by plugging in \( \mathbb{E}[\theta] = s/d \), (b) holds since \( s = \lfloor d/(1 + e^\varepsilon) \rfloor \) (so \( s/d \leq \varepsilon \leq s/d + 1 \)), (c) holds since we only focus on the regime where \( e \leq d - 1 \) (so \( s/d \leq 1 + e^\varepsilon \)), and (d) holds by observing that \( f(x) = \frac{(x-1)^2}{(e^\varepsilon - 1)(x+1)} \) is an increasing function for \( x \geq 1 \) and we have \( e \geq 1 \). Putting things together, we obtain

\[
\frac{m_{ss} - m_{mrc}}{m_{mrc}} = \frac{m_{ss} - (m_{ss} - m_{mrc})}{m_{ss} - (m_{ss} - m_{mrc})} \leq \frac{\sqrt{\ln(2/\beta)/(2N)(e^\varepsilon + 3) + 2\beta}}{0.24 - \left( \frac{\sqrt{\ln(2/\beta)/(2N)(e^\varepsilon + 3) + 2\beta}}{1 + \lambda} \right)} \leq \lambda,
\]

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

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

To ensure (67), let

\[
\beta \leq \frac{0.24\lambda}{4(1 + \lambda)}
\]

\[
N \geq \frac{1}{2} \left( \frac{e^\varepsilon + 3}{0.24\lambda(1 + \lambda) - 2\beta} \right)^2 \ln(2/\beta) = \frac{2(e^\varepsilon + 3)^2(1 + \lambda)^2}{0.24^2\lambda^2} \ln \left( \frac{8(1 + \lambda)}{0.24\lambda} \right).
\]

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

\[ \Box \]

**G.2.2 Relationship between the scaling factors and mean squared errors associated with Subset Selection and MRC simulating Subset Selection**

In the following Proposition, we show that if \( m_{mrc} \) is close to \( m_{ss} \) and \( b_{mrc} \geq b_{ss} \), then the mean squared error associated with MRC simulating Subset Selection is close to the mean squared error associated with Subset Selection.

**Proposition G.1.** Let \( q^{ss}(z|x) \) be the \( \varepsilon \)-LDP Subset Selection mechanism with estimator \( \hat{x}^{ss} \). Let \( q^{mrc}(z|x) \) denote the MRC privatization mechanism simulating Subset Selection with \( N \) candidates and estimator \( \hat{x}^{mrc} \). 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

\[
\mathbb{E}_{q^{mrc}}[\|\hat{x}^{mrc} - x\|_2^2] \leq (1 + 4\lambda + 5\lambda^2 + 2\lambda^3) \mathbb{E}_{q^{ss}}[\|\hat{x}^{ss} - x\|_2^2]
\]
Proof. We have

$$\mathbb{E}_{q_{\text{mrc}}} [||\hat{x}_{\text{mrc}} - x||_2^2] = \sum_{i=1}^{d} \text{Var}(\hat{x}_{i \text{mrc}}) = \left(\frac{1}{m_{\text{mrc}}}\right)^2 \sum_{i} \text{Var}(z_i) = \left(\frac{1}{m_{\text{mrc}}}\right)^2 \sum_{i} q_{i \text{mrc}}(1 - q_{i \text{mrc}}).$$

where (a) follows because $x$ is a constant, (b) follows because $\hat{x}_{\text{mrc}} = (z_K - b_{\text{mrc}})/m_{\text{mrc}}$, and (c) follows because $(z_K)_i \sim \text{Ber}(q_{i \text{mrc}})$. Similarly, we have We have

$$\mathbb{E}_{q_{\text{ss}}} [||\hat{x}_{\text{ss}} - x||_2^2] = \sum_{i=1}^{d} \text{Var}(\hat{x}_{i \text{ss}}) = \left(\frac{1}{m_{\text{ss}}}\right)^2 \sum_{i} \text{Var}(z_i) = \left(\frac{1}{m_{\text{ss}}}\right)^2 \sum_{i} q_{i \text{ss}}(1 - q_{i \text{ss}}).$$

where (a) follows because $x$ is a constant, (b) follows because $\hat{x}_{\text{ss}} = (z - b_{\text{ss}})/m_{\text{ss}}$, and (c) follows because $z_i \sim \text{Ber}(q_{i \text{ss}})$.

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

$$\mathbb{E}_{q_{\text{mrc}}} [||\hat{x}_{\text{mrc}} - x||_2^2] - \mathbb{E}_{q_{\text{ss}}} [||\hat{x}_{\text{ss}} - x||_2^2] = \left(\frac{1}{m_{\text{mrc}}}\right)^2 \sum_{i} q_{i \text{mrc}}(1 - q_{i \text{mrc}}) - \left(\frac{1}{m_{\text{ss}}}\right)^2 \sum_{i} q_{i \text{ss}}(1 - q_{i \text{ss}}) \leq \left(\frac{1}{m_{\text{mrc}}}\right)^2 \sum_{i} (q_{i \text{mrc}} - q_{i \text{ss}})(1 - q_{i \text{ss}}) + \left(\frac{1}{m_{\text{mrc}}}\right)^2 \sum_{i} q_{i \text{ss}}(1 - q_{i \text{ss}}).$$

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

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

Further, we have

$$q_{i \text{mrc}} \leq m_{\text{mrc}}p_i + b_{\text{mrc}} = q_{i \text{ss}} + (m_{\text{mrc}} - m_{\text{ss}})p_i + (b_{\text{mrc}} - b_{\text{ss}}) \geq q_{i \text{ss}} - \lambda \cdot m_{\text{mrc}} \cdot p_i + (b_{\text{mrc}} - b_{\text{ss}}) \geq q_{i \text{ss}} - \lambda \cdot m_{\text{ss}} \cdot p_i \geq (1 - \lambda)q_{i \text{ss}},$$

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

$$\frac{q_{i \text{mrc}}(1 - q_{i \text{mrc}}) - q_{i \text{ss}}(1 - q_{i \text{ss}})}{q_{i \text{ss}}(1 - q_{i \text{ss}})} = \frac{(q_{i \text{ss}} - q_{i \text{mrc}})(q_{i \text{mrc}} + q_{i \text{ss}} - 1)}{q_{i \text{ss}}(1 - q_{i \text{ss}})} \leq \frac{\lambda q_{i \text{ss}}(q_{i \text{ss}} + q_{i \text{mrc}} - 1)}{q_{i \text{ss}}(1 - q_{i \text{ss}})} \leq \frac{\lambda}{1 - q_{i \text{ss}}}$$

where (a) follows from (69) and (b) follows since $q_{i \text{ss}} \leq 1$ and $q_{i \text{mrc}} \leq 1$.

Let us now upper bound $q_{i \text{ss}}$. We have

$$q_{i \text{ss}} = m_{\text{ss}} \cdot p_i + b_{\text{ss}} \leq m_{\text{ss}} + b_{\text{ss}} \leq \frac{\mathbb{E}[\theta]e^{\theta}e}{e^{\mathbb{E}[\theta]} + (1 - \mathbb{E}[\theta])} \leq \frac{1}{2}$$

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

$$\sum_{i} q_{i \text{mrc}}(1 - q_{i \text{mrc}}) - \sum_{i} q_{i \text{ss}}(1 - q_{i \text{ss}}) \leq 2\lambda \sum_{i} q_{i \text{ss}}(1 - q_{i \text{ss}}).$$
Together with (68), we obtain
\[
\left(\frac{1}{m_{\text{mrc}}}\right)^2 \sum_i (q_{i}^{\text{mrc}}(1 - q_{i}^{\text{mrc}}) - q_{i}^{\text{ss}}(1 - q_{i}^{\text{ss}})) \leq \frac{2\lambda(1 + \lambda)^2}{m_{\text{ss}}^2} \sum_i q_{i}^{\text{ss}}(1 - q_{i}^{\text{ss}}).
\]

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

Combining everything, we have
\[
\mathbb{E}_{q^{\text{mrc}}} \left[\|\hat{x}^{\text{mrc}} - x\|_2^2\right] \leq \left(1 + 2\lambda(1 + \lambda)^2 + 2\lambda + \lambda^2\right) \frac{1}{m_{\text{mrc}}} \sum_i q_{i}^{\text{ss}}(1 - q_{i}^{\text{ss}})
\]
\[
= \left(1 + 4\lambda + 5\lambda^2 + 2\lambda^3\right) \mathbb{E}_{q^{\text{ss}}} \left[\|\hat{x}^{\text{ss}} - x\|_2^2\right]
\]

G.2.3 Simulating Subset Selection using Minimal Random Coding

The following Theorem shows that \(q^{\text{mrc}}\) can compress \(q^{\text{ss}}\) to the order of \(\varepsilon\)-bits of communication as well as simulate it in a near-lossless fashion.

**Theorem G.1.** Let \(q^{\text{ss}}(z|x)\) be the \(\varepsilon\)-LDP Subset Selection mechanism with estimator \(\hat{x}^{\text{ss}}\). Let \(q^{\text{mrc}}(z|x)\) denote the MRC privatization mechanism simulating Subset Selection with \(N\) candidates and estimator \(\hat{x}^{\text{mrc}}\). Consider any \(\lambda > 0\). Then,
\[
\mathbb{E}_{q^{\text{mrc}}} \left[\|\hat{x}^{\text{mrc}} - x\|_2^2\right] \leq \left(1 + 4\lambda + 5\lambda^2 + 2\lambda^3\right) \mathbb{E}_{q^{\text{ss}}} \left[\|\hat{x}^{\text{ss}} - x\|_2^2\right]
\]
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 G.1 and Lemma G.2.

G.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 5.1 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 \(\varepsilon\)'s typically considered by LDP mechanisms while only using \((3\varepsilon/\ln 2) + 6\) bits. In comparison with the results from Section 5.1, the results in this section come with an approximate guarantee (\(\delta = 10^{-6}\)) and with a higher number of bits of communication.
H Modified Minimal Random Coding Simulating Subset Selection

In this section, we prove Lemma 5.1 and Theorem 5.1. To prove Theorem 5.1, first, in Appendix H.2.1, we show that when the number of candidates $N$ is exponential in $\varepsilon$, the scaling factor $m_{\text{mmrc}}$ is close to the scaling parameter associated with Subset Selection (i.e., $m_{\text{as}}$). Next, in Appendix H.2.2, we show that if the scaling factor $m_{\text{mmrc}}$ is close to the scaling parameter $m_{\text{as}}$, then the mean squared error associated with MMRC simulating Subset Selection is close to the mean squared error associated with Subset Selection. Finally, in Appendix H.3, we provide some empirical comparisons in addition to the ones in Section 5.1.

H.1 Unbiased Modified Minimal Random Coding simulating Subset Selection

Consider the Subset Selection $\varepsilon$-LDP mechanism $q^{ss}$ described in Section 2 with $s := \left\lceil \frac{d}{1+\varepsilon^2} \right\rceil$. Subset Selection is cap-based mechanism as discussed in Section 3 and Appendix F with $\text{Cap}_x = Z_x$ and $\mathbb{P}_{Z \sim \text{Unif}(Z)}(z \in \text{Cap}_x) = s/d$. Let $\pi_{\text{mmrc}}$ be the distribution and $z_1, z_2, \ldots, z_N$ be the candidates obtained from Algorithm 2 when the reference distribution is $\text{Unif}(Z)$ where $Z$ is as defined in (51). Let $\theta$ denote the fraction of candidates inside $\text{Cap}_x = Z_x$ where $Z_x$ is the set of elements in $Z$ with 1 in the same location as $x$. It is easy to see that $\theta := \frac{1}{N} \text{Binom} \left( N, \frac{1}{2} \right)$. Let $q^i_{\text{mmrc}} = \mathbb{P}(z_i = 1)$ where $z \sim q^\text{mmrc}(\cdot|x)$ i.e., $q^i_{\text{mmrc}} = \mathbb{P}\{ (z_K)_i = 1 \}$ where $K \sim \pi^\text{mmrc}(\cdot)$.

**Lemma H.1.** Let $K \sim \pi^\text{mmrc}(\cdot)$ and $q^i_{\text{mmrc}} = \mathbb{P}\{ (z_K)_i = 1 \}$ for $i \in [d]$. Then,

\[ q^i_{\text{mmrc}} = p_i m_{\text{mmrc}} + b_{\text{mmrc}} \]

where

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

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

**Proof.** Following the proof of Lemma G.1, we compute $\mathbb{P}\{ (z_K)_i = 1 \mid x = i \}$ and $\mathbb{P}\{ (z_K)_i = 1 \mid x = j \}$ separately.

To compute $\mathbb{P}\{ (z_K)_i = 1 \mid x = i \}$, recall that $\theta$ denotes the fraction of candidates that belong inside the $\text{Cap}_x$ i.e., have 1 in the same location as $x$. From Appendix F.2, recall that $c_1(\varepsilon, d) :=$
\[
\frac{e^\varepsilon}{(d-1)e^\varepsilon + (d-1)}, \quad c_2(\varepsilon, d) := \frac{1}{(d-1)e^\varepsilon + (d-1)}.
\]
Further, since \( z_k \) are generated uniformly at random,
\[
\theta \sim \frac{1}{N} \text{Binom} \left( N, \frac{(d-1)}{(d-1)s} \right) = \frac{1}{N} \text{Binom} \left( N, \frac{s}{d} \right),
\]
so we have
\[
\mathbb{P} \{(z_K)_i = 1|x = i\} = \mathbb{P} \{z_K \in \text{Cap}_x|x = i\} \overset{(a)}{=} \mathbb{E} \left[ \mathbb{P} \{z_K \in \text{Cap}_x|x = i, \theta\} \right] = \mathbb{E} \left[ \frac{e^\varepsilon}{e^\varepsilon \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{I} (\theta \leq \mathbb{E}[\theta]) + \frac{e^\varepsilon \mathbb{E}[\theta] + \theta - \mathbb{E}[\theta]}{e^\varepsilon \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{I} (\theta > \mathbb{E}[\theta]) \right]
\]
(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 \( \mathbb{P} \{(z_K)_i = 1|x = j\} \), we decompose it into
\[
\mathbb{P} \{(z_K)_i = 1|x = j\} = \mathbb{P} \{(z_K)_i = 1, (z_K)_j = 1|x = j\} + \mathbb{P} \{(z_K)_i = 1, (z_K)_j = 0|x = j\}\]
for any \( j \neq i \) and calculate each of the terms separately.

As before, let \( \theta \) denotes the fraction of candidates that belong inside the \( \text{Cap}_x \) i.e., have 1 in the same location as \( x \). Further, let \( \bar{\theta} \) denotes the fraction of candidates that belong inside the \( \text{Cap}_x \) i.e., have 1 in the same location as \( x \) as well as have 1 in the \( j^{th} \) location. Since \( z_k \) are generated uniformly at random,
\[
\bar{\theta} \sim \frac{1}{N} \text{Binom} \left( N \theta, \frac{(d-1)}{(d-1)s} \right) = \frac{1}{N} \text{Binom} \left( N \theta, \frac{s-1}{d-1} \right),
\]
so we have
\[
\mathbb{P} \{(z_K)_i = 1, (z_K)_j = 1|x = j\} \overset{(a)}{=} \mathbb{E}_{\bar{\theta}} \left[ \mathbb{E}_{\theta} \left[ \mathbb{P} \{(z_K)_i = 1, (z_K)_j = 1|x = j, \bar{\theta}, \theta\} \right] \right] = \mathbb{E}_{\bar{\theta}} \left[ \frac{e^\varepsilon \theta}{e^\varepsilon \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{I} (\theta \leq \mathbb{E}[\theta]) + \frac{e^\varepsilon \mathbb{E}[\theta] + \theta - \mathbb{E}[\theta]}{e^\varepsilon \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{I} (\theta > \mathbb{E}[\theta]) \right]
\]
(75)
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 \( \mathbb{P} \{(z_K)_i = 1, (z_K)_j = 0|x = j\} \), let \( \bar{\theta} \) denote the fraction of candidates that belong inside the \( \text{Cap}_x \) i.e., have 1 in the same location as \( x \) as well as have 0 in the \( j^{th} \) location. Since \( z_k \) are generated uniformly at random,
\[
\bar{\theta} \sim \frac{1}{N} \text{Binom} \left( N(1 - \theta), \frac{(d-1)}{(d-1)s} \right) = \frac{1}{N} \text{Binom} \left( N(1 - \theta), \frac{s}{d} \right),
\]
so we have
\[
\mathbb{P} \{(z_K)_i = 1, (z_K)_j = 0|x = j\} \overset{(a)}{=} \mathbb{E}_{\bar{\theta}} \left[ \mathbb{E}_{\theta} \left[ \mathbb{P} \{(z_K)_i = 1, (z_K)_j = 0|x = j, \bar{\theta}, \theta\} \right] \right] = \mathbb{E}_{\bar{\theta}} \left[ \frac{\bar{\theta}}{e^\varepsilon \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{I} (\theta \leq \mathbb{E}[\theta]) + (1 - \mathbb{E}[\theta]) \left( \frac{e^\varepsilon \mathbb{E}[\theta] + \theta - \mathbb{E}[\theta]}{e^\varepsilon \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \bar{\theta} \cdot \mathbb{I} (\theta > \mathbb{E}[\theta]) \right) \right]
\]
(76)
\[
\frac{s}{d-1} \mathbb{E} \left[ \frac{(1-\theta)}{e^{\theta} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{1}(\theta > \mathbb{E}[\theta]) + \frac{(1 - \mathbb{E}[\theta]) + (\mathbb{E}[\theta] - \theta) e^\varepsilon}{e^{\theta} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{1}(\theta \leq \mathbb{E}[\theta]) \right].
\]

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

\[
P\{(z_K)_i = 1|x = j\} = \frac{1}{d-1} \left( s - \mathbb{E} \left[ \frac{e^\varepsilon \theta}{e^{\theta} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{1}(\theta \leq \mathbb{E}[\theta]) + \frac{e^{\theta} \mathbb{E}[\theta] + (1 - \theta) e^\varepsilon}{e^{\theta} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{1}(\theta > \mathbb{E}[\theta]) \right] \right).
\]

Combining everything, we have

\[
q_i^{mrrc} = \mathbb{P}\{(z_K)_i = 1\} = p_i \times \left[ \mathbb{P}\{(z_K)_i = 1|x = i\} - \mathbb{P}\{(z_K)_i = 1|x = j\} \right] + \mathbb{P}\{(z_K)_i = 1|x = j\} \]

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

**Lemma 5.1.** Let \( \hat{x}^{mrrc} \) be the estimator of the MMRRC mechanism simulating Subset Selection as defined above. Then, \( \mathbb{E}[\hat{x}^{mrrc}] = x \).

**Proof.** Given Lemma H.1, the proof follows from the proof of Lemma G.1.

### H.2 Utility of Modified Minimal Random Coding simulating Subset Selection

#### H.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 \( \varepsilon \), then the scaling parameters associated with Subset Selection and the MMRC scheme simulating Subset Selection are close.

**Lemma H.2.** Let \( N \) denote the number of candidates used in the MMRC scheme. Let \( K \sim \pi^{mrrc} \) where \( \pi^{mrrc} \) 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 \( m_{mrrc} \) and \( b_{mrrc} \) associated with the MMRC scheme simulating Subset Selection are such that

\[
m_{ss} - m_{mrrc} \leq \lambda \cdot m_{mrrc}
\]

and \( b_{ss} \leq b_{mrrc} \) 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 G.2. We only show the key steps here.

From (73) and (63), we have

\[
b_{mrrc} - b_{ss} = \frac{1}{d-1} \cdot \frac{1}{e^{\theta} \mathbb{E}[\theta] + (1 - \mathbb{E}[\theta])} \cdot \mathbb{E} \left[ e^{\varepsilon}(\mathbb{E}[\theta] - \theta) \cdot \mathbb{1}(\theta \leq \mathbb{E}[\theta]) + (\mathbb{E}[\theta] - \theta) \cdot \mathbb{1}(\theta > \mathbb{E}[\theta]) \right].
\]
\[
\begin{align*}
\text{(a)} \quad & \frac{1}{d-1} \cdot \frac{1}{e^{cE[|\theta|]} + (1 - E[|\theta|])} \cdot E[(E[|\theta|] - \theta) \cdot \mathbb{1}(\theta \leq E[|\theta|]) + (E[|\theta|] - \theta) \cdot \mathbb{1}(\theta > E[|\theta|])] \\
= & \frac{1}{d-1} \cdot \frac{1}{e^{cE[|\theta|]} + (1 - E[|\theta|])} \cdot E[(E[|\theta|] - \theta)] = 0.
\end{align*}
\]

where (a) follows because \(e^c \geq 1\). From (72) and (62), we have

\[
m_{ss} - m_{mmrc} = \frac{d}{d-1} \cdot \frac{1}{e^{cE[|\theta|]} + (1 - E[|\theta|])} \cdot E[e^c(E[|\theta|] - \theta) \cdot \mathbb{1}(\theta \leq E[|\theta|]) + (E[|\theta|] - \theta) \cdot \mathbb{1}(\theta > E[|\theta|])]
\]

\[
\leq \frac{d}{d-1} \cdot \frac{1}{e^{cE[|\theta|]} + (1 - E[|\theta|])} \cdot E[e^c(E[|\theta|] - \theta) \cdot \mathbb{1}(\theta \leq E[|\theta|])]
\]

\[
\leq \frac{2}{e^{cE[|\theta|]} + (1 - E[|\theta|])} \cdot E[e^c(E[|\theta|] - \theta) \cdot \mathbb{1}(\theta \leq E[|\theta|])]
\]

where (a) holds since \(d \geq 2\). Next, we condition on the event \(E := \{|E[|\theta|] - \theta| \leq \sqrt{\frac{\ln(2/\beta)}{2N}}\}\), which has probability \(P_{\theta} \{E\} \geq 1 - \beta\) by Hoeffding’s inequality. We continue to upper bound (79):

\[
m_{ss} - m_{mmrc} = 2P \{E\} \cdot E \left[ \frac{e^c(E[|\theta|] - \theta) \cdot \mathbb{1}(\theta \leq E[|\theta|])}{e^{cE[|\theta|]} + (1 - E[|\theta|])} \right] + 2P \{E^c\} \cdot E \left[ \frac{e^c(E[|\theta|] - \theta) \cdot \mathbb{1}(\theta \leq E[|\theta|])}{e^{cE[|\theta|]} + (1 - E[|\theta|])} \right] \]

\[
\leq 2 \left( E \left[ \frac{e^c(E[|\theta|] - \theta) \cdot \mathbb{1}(\theta \leq E[|\theta|])}{e^{cE[|\theta|]} + (1 - E[|\theta|])} \right] + \beta \right)
\]

\[
\leq (1 + e^c) \sqrt{\frac{\ln(2/\beta)}{2N}} + 2\beta
\]

where (a) holds since

\[
\frac{e^c(E[|\theta|] - \theta) \cdot \mathbb{1}(\theta \leq E[|\theta|])}{e^{cE[|\theta|]} + (1 - E[|\theta|])} = \frac{e^cE[|\theta|] \cdot \mathbb{1}(\theta \leq E[|\theta|])}{e^{cE[|\theta|]} + (1 - E[|\theta|])} - \frac{e^{c\theta} \cdot \mathbb{1}(\theta \leq E[|\theta|])}{e^{cE[|\theta|]} + (1 - E[|\theta|])} \leq 1,
\]

and (b) holds since \(E[|\theta|] = s/d \geq 1/(1 + e^c)\).

The rest of the proof is similar to the proof of Lemma G.2.

H.2.2 Relationship between the scaling factors and mean squared errors associated with Subset Selection and MMRC simulating Subset Selection

In the following Proposition, we show that if \(m_{mmrc}\) is close to \(m_{ss}\) and \(b_{mmrc} \geq b_{ss}\), then the mean squared error associated with MMRC simulating Subset Selection is close to the mean squared error associated with Subset Selection.

**Proposition H.1.** Let \(q^{ss}(z|x)\) be the \(\varepsilon\)-LDP Subset Selection mechanism with estimator \(\hat{x}^{ss}\). Let \(q^{mmrc}(z|x)\) denote the MMRC privatization mechanism simulating Subset Selection with \(N\) candidates and estimator \(\hat{x}^{mmrc}\). 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 \(m_{ps} - m_{mmrc} \leq \lambda \cdot m_{mmrc}\) and \(b_{mmrc} \geq b_{ss}\), then

\[
E_{q^{mmrc}}[\|\hat{x}^{mmrc} - x\|^2] \leq (1 + 4\lambda + 5\lambda^2 + 2\lambda^3) E_{q^{ss}}[\|\hat{x}^{ss} - x\|^2]
\]

**Proof.** The proof is similar to the proof of Proposition G.1.
H.2.3 Simulating Subset Selection using Modified Minimal Random Coding

Now, we provide a proof of Theorem 5.1.

**Theorem 5.1.** Let $q^{ss}(z|x)$ be the $\epsilon$-LDP Subset Selection mechanism with parameters $d$ and $s = \lceil \frac{d}{1 + \epsilon^2} \rceil$ and estimator $\hat{x}^{ss}$. Let $q^{mmrc}(z|x)$ denote the MMRC privatization mechanism simulating Subset Selection with $N$ candidates and estimator $\hat{x}^{mmrc}$ as defined above. Consider any $\lambda > 0$. Then,

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

as long as

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

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

H.3 Additional Empirical Comparisons

In Section 5.1, we empirically demonstrated the privacy-accuracy-communication tradeoffs of MMRC simulating Subset Selection against Subset Selection and RHR in terms of $\ell_2$ error vs #bits and $\ell_2$ error vs $\epsilon$ (see Figure 2). In this section, we provide comparisons between these methods in terms of $\ell_2$ error vs $d$ (see Figure 6 (left)) and $\ell_2$ error vs $n$ (see Figure 6 (right)) for a fixed $\epsilon$ (=6) and a fixed #bits (=14). As before, RHR uses #bits = $\epsilon$ for both because it leads to a poor performance if #bits > $\epsilon$.

![Figure 6](image_url)

Figure 6: Comparing Subset Selection, MMRC simulating Subset Selection and RHR for frequency estimation with $\epsilon = 6$ and #bits = 14. **Left:** $\ell_2$ error vs $d$ for $n = 5000$. **Right:** $\ell_2$ error vs $n$ for $d = 500$. 

50