A Proof of Lemma 3.1

Proof. For \( n \)-user mean estimation protocol \((f, \mathcal{A}, P_{uf})\), following the notation and steps from Proof of Lemma 3.1, we define the marginalized output

\[
\hat{g}_i(m_i, U_i; v^n) = E_{(m_j, U_j)_{j \neq i}} \left[ n\mathcal{A}(\{m_j, U_j\}_{j=1}^n) \right] \left| f_i(v_i, U_i) = m_i, U_i, v^{n\setminus i} \right].
\]  

(35)

Then, we define the user-specific decoder by averaging \( g_i(m_i, U_i; v^n) \) with respect to i.i.d. uniform \( P_{\text{unif}} \):

\[
g_i(m_i, U_i) = E_{v^{n\setminus i} \sim P_{\text{unif}}} [\hat{g}_i(m_i, U_i; v^n)]
\]

(36)

where \( v^{n\setminus i} \) indicates the \( v^n \) vector except \( v_i \). Due to the symmetry of \( P_{\text{unif}} \), it is clear that \( g_i \) is unbiased. We also define

\[
\hat{R}_{\leq i}(\{v_j, m_j, U_j\}_{j=1}^n) = E_{v_j \sim P_{\text{unif}}, j \neq i} \left[ n\mathcal{A}(\{m_j, U_j\}_{j=1}^n) - \sum_{j=1}^{i} v_j \left| \{v_j, m_j, U_j\}_{j=1}^n \right. \right]
\]

(37)

Consider an average error where \( v_1, \ldots, v_n \) are drawn i.i.d. uniformly on the sphere \( S^{d-1} \).

\[
E_{(v_j, m_j, U_j)_{j=1}^n} \left[ \left\| n\mathcal{A}(\{m_j, U_j\}_{j=1}^n) - \sum_{j=1}^n v_j \right\|^2 \right]
\]

(38)

\[
= E_{(v_j, m_j, U_j)_{j=1}^n} \left[ \left\| \hat{R}_{\leq n}(\{v_j, m_j, U_j\}_{j=1}^n) \right\|^2 \right]
\]

(39)

\[
= E_{(v_j, m_j, U_j)_{j=1}^n} \left[ \left\| \hat{R}_{\leq n}(\{v_j, m_j, U_j\}_{j=1}^n) - \hat{R}_{\leq n-1}(\{v_j, m_j, U_j\}_{j=1}^{n-1}) + \hat{R}_{\leq n-1}(\{v_j, m_j, U_j\}_{j=1}^{n-1}) \right\|^2 \right]
\]

\[
= \sum_{i=1}^{n} \left[ \left\| \hat{R}_{\leq i}(\{v_j, m_j, U_j\}_{j=1}^n) - \hat{R}_{\leq i-1}(\{v_j, m_j, U_j\}_{j=1}^{i-1}) \right\|^2 \right]
\]

(40)

\[
\geq \sum_{i=1}^{n} E_{m_i, U_i} \left[ \left\| \hat{R}_{\leq i}(\{v_j, m_j, U_j\}_{j=1}^n) - \hat{R}_{\leq i-1}(\{v_j, m_j, U_j\}_{j=1}^{i-1}) \right\|^2 \right]
\]

(41)

\[
= \sum_{i=1}^{n} E_{m_i, U_i} \left[ \|g_i(m_i, U_i) - v_i\|^2 \right]
\]

(42)

Then, we need to show the same inequality for the worst-case error.

\[
\sup_{v_1, \ldots, v_n} E_{(m_j, U_j)_{j=1}^n} \left[ \left\| n\mathcal{A}(\{m_j, U_j\}_{j=1}^n) - \sum_{j=1}^n v_j \right\|^2 \right]
\]

(43)

\[
\geq E_{(v_j, m_j, U_j)_{j=1}^n} \left[ \left\| n\mathcal{A}(\{m_j, U_j\}_{j=1}^n) - \sum_{j=1}^n v_j \right\|^2 \right]
\]

(44)

\[
= \sum_{i=1}^{n} E_{v_i, m_i, U_i} \left[ \|g_i(m_i, U_i) - v_i\|^2 \right]
\]

(45)

\[
= \sum_{i=1}^{n} \sup_{v_i} E_{m_i, U_i} \left[ \|g_i(m_i, U_i) - v_i\|^2 \right]
\]

(46)
where the last equality \((46)\) is from Lemma 3.2, Lemma 3.4 and Lemma 3.5. Thus, the user-specific decoder achieves lower MSE:

\[
\text{Err}_n(f, A, P_{U^n}) \leq \frac{1}{n} \sum_{i=1}^n \text{Err}_1(f_i, g_i, P_{U_i}).
\]

(47)

Since we keep random encoder \(f_i\) the same, the canonical protocol with \(g_i\) also satisfies \(\varepsilon\)-LDP constraint. This concludes the proof.

\[\square\]

**B Proof of Lemma 3.2**

**Proof.** Let \(\tilde{U}_m = g(m, U)\) for all \(1 \leq m \leq M\). Without loss of generality \(g(\cdot, U)\) is one-to-one, i.e., \(\{ u : \tilde{u}_m = g(m, u) \text{ for all } m \}\) has at most one element (with probability 1), and \(u = g^{-1}(\tilde{u}^M)\) is well-defined. Then, we define a randomized encoder \(f_0(v, \tilde{U}^M)\) that satisfies

\[
Q_{f_0}(m|v, \tilde{u}^M) = Q_f(m|v, g^{-1}(\tilde{u}^M)).
\]

(48)

It is clear that \(f_0\) satisfies \(\varepsilon\)-LDP constraint. Then,

\[
D(v, f_0, g^+, P_{\tilde{U}^M}) = \mathbb{E}_{f_0, P_{U^M}} \left[ ||g^+(f_0(v, \tilde{U}^M), \tilde{U}^M) - v||^2 \right]
\]

(49)

\[
= \mathbb{E}_{f_0, P_{U^M}} \left[ \sum_{m=1}^M Q_{f_0}(m|v, \tilde{U}^M)||\tilde{U}_m - v||^2 \right]
\]

(50)

\[
= \mathbb{E}_{f, P_U} \left[ \sum_{m=1}^M Q_f(m|v, U)||g(m, U) - v||^2 \right]
\]

(51)

\[
= \mathbb{E}_{f, P_U} \left[ ||g(f(v, U), U) - v||^2 \right]
\]

(52)

\[
= D(v, f, g, P_U).
\]

(53)

We also need to show that the composition of the new randomizer \(f_0\) and selector \(g^+\) is unbiased.

\[
\mathbb{E}_{P_{\tilde{U}^M}} \left[ g^+(f_0(v, \tilde{U}^M), \tilde{U}^M) \right] = \mathbb{E}_{f_0, P_{U^M}} \left[ \sum_{m=1}^M Q_{f_0}(m|v, \tilde{U}^M)\tilde{U}_m \right]
\]

(54)

\[
= \mathbb{E}_{f, P_U} \left[ \sum_{m=1}^M Q_f(m|v, U)g(m, U) \right]
\]

(55)

\[
= \mathbb{E}_{f, P_U} \left[ g(f(v, U), U) \right]
\]

(56)

\[
= v.
\]

(57)

Finally, \(Q_{f_0}(m|v, \tilde{u}^M)\) is a valid transition probability, since

\[
\sum_{m=1}^M Q_{f_0}(m|v, \tilde{u}^M) = \sum_{m=1}^M Q_f(m|v, g^{-1}(\tilde{u}^M)) = 1
\]

(58)

for all \(\tilde{u}^M\). This concludes the proof.

\[\square\]

**C Proof of Lemma 3.4**

**Proof.** Let \(A\) be a uniformly random orthogonal matrix and \(\tilde{U}^M = AU^M\). We further let \(f_1\) be a randomized encoder that satisfies

\[
Q_{f_1}(m|v, \tilde{U}^M) = \mathbb{E}_A \left[ Q_f(m|Av, AU^M)|\tilde{U}^M \right].
\]

(59)

Then, \(Q_{f_1}\) is a valid probability since

\[
\sum_{m=1}^M Q_{f_1}(m|v, \tilde{U}^M) = \mathbb{E}_A \left[ \sum_{m=1}^M Q_f(m|Av, AU^M)|\tilde{U}^M \right] = 1.
\]

(60)

Also, we have

\[
\frac{Q_{f_1}(m|v, \tilde{U}^M)}{Q_{f_1}(m|v', \tilde{U}^M)} = \mathbb{E}_A \left[ \frac{Q_f(m|Av, AU^M)|\tilde{U}^M}{Q_f(m|Av', AU^M)|\tilde{U}^M} \right]
\]

(61)
The key step is that the original encoder for all $\forall v, v'$.

Similarly to the previous proofs, $D$ Proof of Lemma 3.5

For $v, v' \in S^{d-1}$, let $A_0$ be an orthonormal matrix such that $v' = A_0v$. Let $f_2$ be a randomized encoder such that

$$f_2(v, U^M) = f(Av, AU^M)$$

for uniform random orthonormal matrix. Then,

$$Q_{f_2}(m|v, U^M) = E_A [Q_f(m|Av, AU^M)].$$

(82)

Similarly to the previous proofs, $Q_{f_2}$ is a well-defined probability distribution, and $f_2$ is unbiased as well as $\varepsilon$-LDP. Since $P_{U, M}$ is rotationally symmetric and $f_2$ is also randomized via the uniform random orthogonal matrix, we have

$$D(v', f_2, P_{U, M}) = D(A_0v, f_2, P_{U, M}) = D(v, f_2, P_{U, M}).$$

(83)

Compared to a given randomizer $f$, we have

$$\text{Err}(f, P_{U, M}) \geq E_A [D(Av, f, P_{U, M})],$$

(84)

for all $v \in S^{d-1}$. This concludes the proof.
E Proof of Theorem 3.6

Proof. The rotationally symmetric simplex codebook with normalization constant $r$ is $(rA_s, \ldots, rA_{sM})$. Let $f$ be the unbiased encoder satisfying $\varepsilon$-LDP. Let $Q_{\text{max}} = \max Q_f(m|v, rA_{sM})$ and $Q_{\text{min}} = \min Q_f(m|v, rA_{sM})$, our objective is to demonstrate that $Q_{\text{max}}$ is less than or equal to $e\varepsilon Q_{\text{min}}$. We will employ a proof by contradiction to establish this. Suppose $Q_f(m_1|v_1, rA_{sM}) > e\varepsilon Q_f(m_2|v_2, rA_{sM})$ for some $m_1, v_1, A_1, m_2, v_2, A_2$. Let $\hat{A}$ be the row switching matrix where $\hat{A}A_1s_{m_1} = rA_1s_{m_2}$ and $\hat{A}A_1s_{m_2} = rA_1s_{m_1}$, then we have

$$Q_f(m_1|v_1, rA_{sM}) = Q_f(m_1|\hat{A}v_1, r\hat{A}A_1s_{m_1}).$$

We further let $A'$ be an orthogonal matrix such that $A'\hat{A}A_1 = A_2$, then

$$Q_f(m_2|\hat{A}v_1, r\hat{A}A_1s_{m_1}) = Q_f(m_2|A'\hat{A}v_1, rA'\hat{A}A_1s_{m_1}) = Q_f(m_2|A'\hat{A}v_1, rA_2s_{m_1}).$$

If we let $v'_1 = A'\hat{A}v_1$, then

$$Q_f(m_2|v'_1, rA_2s_{m_1}) = Q_f(m_1|v_1, rA_1s_{m_1}),$$

which contradicts the $\varepsilon$-LDP constraint.

For an unbiased encoder, the error is

$$\mathbb{E}_{U,M} \left[ \sum_{m=1}^{M} \|U_m - v\|^2 Q_f(m|v, U^M) \right] = \mathbb{E}_{U,M} \left[ \sum_{m=1}^{M} \|U_m\|^2 Q_f(m|v, U^M) \right] - 1 = r^2 - 1.$$ (92)

Thus, we need to find $r$ that minimizes the error.

On the other hand, the encoder needs to satisfy unbiasedness. Without loss of generality, we assume $v = e_1$, then we need

$$\mathbb{E}_{A} \left[ \sum_{m=1}^{M} rA_m Q_f(m|e_1, rA_{sM}) \right] = e_1,$$ (93)

where the expectation is with respect to the random orthonormal matrix $A$. If we focus on the first index of the vector, then

$$r \times \mathbb{E}_{a} \left[ \sum_{m=1}^{M} a^Ts_m Q_f(m|e_1, rA_{sM}) \right] = 1,$$ (94)

where $a^T = (a_1, \ldots, a_d)$ is the first row of $A$ and has uniform distribution on the sphere $S^{d-1}$. Thus, it is clear that assigning higher probability (close to $Q_{\text{max}}$) to the larger $a^Ts_m$.

If $Q_{\text{max}}$ is strictly smaller than $e\varepsilon Q_{\text{min}}$, then we can always scale up the larger probabilities and scale down the lower probabilities to keep the probability sum to one (while decreasing the error). Hence, we can assume that $Q_{\text{min}} = q_0$ and $Q_{\text{max}} = e\varepsilon q_0$ for some $1 > q_0 > 0$.

Now, let $k$ be such that

$$(M - |k| - 1)q_0 + q_i + |k|e\varepsilon q_0 = 1,$$ (95)

where $q_i$ is an intermediate value such that $q_i \in [q_0, e\varepsilon q_0]$. Then, the optimal strategy is clear: (i) assign $e\varepsilon q_0$ to $|k|$-th closest codewords $s_{m_i}$'s, (ii) assign $q_i$ to the $(|k| + 1)$-th closest codeword, and (iii) assign $q_0$ to the remaining codewords. This implies that the $k$-closest coding is optimal.

F Proof of Lemma 3.7

Proof. Following (28) with $U_m = As_m$ and $v = e_1$, we have

$$r_k \frac{e\varepsilon - 1}{ke\varepsilon + (M - k)} \mathbb{E} \left[ \sum_{m \in T_k(e_1, A^s)} A \cdot s_m \right]$$
where \( (a) \) holds due to the spherical symmetry of \( \mathbf{a} \). By focusing on the first coordinate of the above equation and observing that \( (e_1, A_\mathcal{S}) = (a, s_m) \) where \( a \) is the first row of the rotation matrix \( A \), we must have

\[
\rho_k \cdot \frac{e^x - 1}{ke^x + (M - k)} \mathbb{E}_{\mathbf{a} \sim \text{unif}(\mathbb{S}^{d-1})} \left[ \sum_{m \in \text{arg max}_k(\{ (e_1, A_s) \})} A \cdot s_m \right] = e_1.
\]

Next, observe that by definition,

\[
s_m = \frac{M}{\sqrt{M(M-1)}} a_m - \frac{1}{\sqrt{M(M-1)}} \mathbf{1}_M,
\]

where \( \mathbf{1}_M = (1, 1, \ldots, 1, 0, \ldots, 0) \in \{0, 1\}^d \) (that is, \( \mathbf{1}_M)_m = 1 \{ m \leq M \} \). Therefore,

\[
\langle a, s_m \rangle = \frac{M}{\sqrt{M(M-1)}} a_m - \frac{1}{\sqrt{M(M-1)}} \langle a, \mathbf{1}_M \rangle,
\]

and hence plugging in (96) yields

\[
\rho_k \cdot \frac{e^x - 1}{ke^x + (M - k)} \mathbb{E}_{\mathbf{a} \sim \text{unif}(\mathbb{S}^{d-1})} \left[ \sum_{m \in \text{Top}_k(\{ (a, s_m) \})} \langle a, s_m \rangle \right] = r_k \cdot \frac{e^x - 1}{ke^x + (M - k)} \frac{M}{\sqrt{M(M-1)}} \mathbb{E}_{\mathbf{a} \sim \text{unif}(\mathbb{S}^{d-1})} \left[ \sum_{i=1}^k a_{(i)M} - \frac{k}{M} \langle a, \mathbf{1}_M \rangle \right]
\]

\[
= r_k \cdot \frac{e^x - 1}{ke^x + (M - k)} \frac{M}{\sqrt{M(M-1)}} \mathbb{E}_{\mathbf{a} \sim \text{unif}(\mathbb{S}^{d-1})} \left[ \sum_{i=1}^k a_{(i)M} \right],
\]

where (1) \( a_{(i)M} \) denotes the \( i \)-th largest entry of the first \( M \) coordinates of \( a \) and (2) the last equality holds since \( a \) is uniformly distributed over \( \mathbb{S}^{d-1} \).

\[\blacksquare\]

G  Proof of Lemma 3.8

\[\text{Proof.}\] First of all, observe that

\[
\mathbb{E}_{\mathbf{a} \sim \text{unif}(\mathbb{S}^{d-1})} \left[ \sum_{i=1}^k a_{(i)M} \right] = \mathbb{E}_{\mathbf{a} \sim \text{unif}(\mathbb{S}^{d-1})} \left[ \mathbb{E} \left[ \sum_{i=1}^k a_{(i)M} \right] \right]
\]

\[
= (a) \mathbb{E}_{\mathbf{a} \sim \text{unif}(\mathbb{S}^{d-1})} \left[ \sum_{i=1}^k a_{(i)M}^2 \cdot \mathbb{E}_{\mathbf{a}_1', \ldots, \mathbf{a}_M'} \sim \text{unif}(\mathbb{S}^{M-1})} \left[ \sum_{i=1}^k a_{(i)M}^2 \right] \right]
\]

\[
= \mathbb{E}_{\mathbf{a} \sim \text{unif}(\mathbb{S}^{d-1})} \left[ \sum_{i=1}^k a_{(i)M}^2 \right] \cdot \mathbb{E}_{\mathbf{a}_1', \ldots, \mathbf{a}_M'} \sim \text{unif}(\mathbb{S}^{M-1})} \left[ \sum_{i=1}^k a_{(i)M}^2 \right],
\]

where (a) holds due to the spherical symmetry of \( \mathbf{a} \). Next, we bound (i) and (ii) separately.
Claim G.1 (Bounding (i)). For any $d \geq M > 2$, it holds that
\[
\sqrt{\frac{M - 2}{d - 2}} \leq \mathbb{E}_{a \sim \text{unif}(S^{d-1})} \left[ \sqrt{\sum_{i=1}^{M} a_i^2} \right] \leq \sqrt{\frac{M}{d - 2}}.
\] (97)

Proof of Claim G.1. Observe that when $a$ is distributed uniformly over $S^{d-1}$, it holds that
\[
(a_1, a_2, \ldots, a_d) \overset{d}{=} \left( \frac{Z_1}{\sqrt{\sum_{i=1}^{d} Z_i^2}}, \frac{Z_2}{\sqrt{\sum_{i=1}^{d} Z_i^2}}, \ldots, \frac{Z_d}{\sqrt{\sum_{i=1}^{d} Z_i^2}} \right),
\]
where $A \overset{d}{=} B$ denotes $A$ and $B$ have the same distribution, and $Z_1, \ldots, Z_d \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$. As a result, we must have
\[
\mathbb{E}_{a \sim \text{unif}(S^{d-1})} \left[ \sqrt{\sum_{i=1}^{M} a_i^2} \right] = \mathbb{E}_{Z_1, \ldots, Z_M \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)} \left[ \sqrt{ \frac{\sum_{i=1}^{M} Z_i^2}{\sum_{i=1}^{M} Z_i^2 + \sum_{i' = M+1}^{d} Z_{i'}^2} } \right].
\]
By Jensen’s inequality, it holds that
\[
\mathbb{E}_{Z_1, \ldots, Z_M \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)} \left[ \sqrt{ \frac{\sum_{i=1}^{M} Z_i^2}{\sum_{i=1}^{M} Z_i^2 + \sum_{i' = M+1}^{d} Z_{i'}^2} } \right] \geq \frac{1}{\sqrt{\sum_{i=1}^{M} Z_i^2 + \sum_{i' = M+1}^{d} Z_{i'}^2}} \mathbb{E}_{Z_1, \ldots, Z_M \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)} \left[ \frac{1}{1 + \frac{\sum_{i' = M+1}^{d} Z_{i'}^2}{\sum_{i=1}^{M} Z_i^2}} \right]
\]
\[
= \frac{1}{1 + \frac{d - M}{M - 2}} \cdot \frac{\sqrt{M - 2}}{d - 2},
\]
where (a) holds since $x \mapsto \sqrt{1/(1 + x)}$ is a convex mapping for $x > 0$, and (b) holds due to the fact that $\sum_i Z_i^2$ follows from a $\chi^2$ distribution and that the ratio of two independent $\chi^2$ random variables follows an $F$-distribution.

On the other hand, it also holds that
\[
\mathbb{E}_{Z_1, \ldots, Z_M \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)} \left[ \sqrt{ \frac{\sum_{i=1}^{M} Z_i^2}{\sum_{i=1}^{M} Z_i^2 + \sum_{i' = M+1}^{d} Z_{i'}^2} } \right] \leq \frac{1}{1 + \frac{\sum_{i=1}^{M} Z_i^2}{\sum_{i=1}^{M} Z_i^2 + \sum_{i' = M+1}^{d} Z_{i'}^2}} \mathbb{E}_{Z_1, \ldots, Z_M \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)} \left[ \frac{1}{1 + \frac{\sum_{i=1}^{M} Z_i^2}{\sum_{i=1}^{M} Z_i^2 + \sum_{i' = M+1}^{d} Z_{i'}^2}} \right]
\]
\[
= 1 - \mathbb{E}_{Z_1, \ldots, Z_M \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)} \left[ \frac{1}{1 + \frac{\sum_{i=1}^{M} Z_i^2}{\sum_{i=1}^{M} Z_i^2 + \sum_{i' = M+1}^{d} Z_{i'}^2}} \right]
\]
where (a) holds since $\sqrt{\cdot}$ is concave, (b) holds since $x \mapsto \frac{1}{1+x}$ is convex, and (c) again is due to the fact that the ratio of two independent $\chi^2$ random variables follows an $F$-distribution.

**Claim G.2** (Bounding (ii)). As long as

- $k \geq 400 \cdot \log 10$,
- $\log (M/k) \geq \left( \frac{10^3 \pi \log 2}{9} \right)^2$,

it holds that

$$\sqrt{\frac{k \log \left( \frac{M}{k} \right)}{24 \pi \log 2M}} \leq \mathbb{E}_{(a_1', \ldots, a_M') \sim \text{unif}(\mathbb{S}^{M-1})} \left[ \sum_{i=1}^{k} a_i' \right] \leq \sqrt{\frac{4k \log M}{M}}. \quad (98)$$

**Proof of Claim G.2** We start by re-writing $a'$:

$$(a_1', a_2', \ldots, a_M') \overset{d}{=} \left( \frac{Z_1}{\sqrt{\sum_{i=1}^{M} Z_i^2}}, \frac{Z_2}{\sqrt{\sum_{i=1}^{M} Z_i^2}}, \ldots, \frac{Z_M}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \right).$$

This yields that

$$(a_1'(1), a_2'(2), \ldots, a_k'(k)) \overset{d}{=} \left( \frac{Z_1(1)}{\sqrt{\sum_{i=1}^{M} Z_i^2}}, \frac{Z_2(2)}{\sqrt{\sum_{i=1}^{M} Z_i^2}}, \ldots, \frac{Z_k(k)}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \right),$$

and hence

$$\mathbb{E}_{(a_1', \ldots, a_M') \sim \text{unif}(\mathbb{S}^{M-1})} \left[ \sum_{i=1}^{k} a_i' \right] = \mathbb{E}_{Z_1, \ldots, Z_M \sim \text{N}(0,1)} \left[ \frac{1}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \sum_{i=1}^{k} Z_i \right].$$

**Upper bound.** To upper bound the above, observe that

$$\mathbb{E}_{Z_1, \ldots, Z_M \sim \text{N}(0,1)} \left[ \frac{1}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \sum_{i=1}^{k} Z_i \right] \leq k \mathbb{E}_{Z_1, \ldots, Z_M \sim \text{N}(0,1)} \left[ \frac{1}{\sqrt{\sum_{i=1}^{M} Z_i^2}} Z_1 \right].$$

Let $\mathcal{E}_1 := \left\{ (Z_1, \ldots, Z_M) \big| \sum_{i=1}^{M} Z_i^2 \leq M(1-\gamma) \right\}$ where $\gamma > 0$ will be optimized later. Then it holds that

$$\Pr \{ \mathcal{E}_1 \} \leq e^{-\frac{M\gamma^2}{4}}. \quad (99)$$

On the other hand, the Borell-TIS inequality ensures

$$\Pr \left\{ \left| Z_1 - \mathbb{E} [Z_1] \right| > \xi \right\} \leq 2e^{-\frac{\xi^2}{2\sigma^2}}, \quad (100)$$

17
where \( Z_i \sim \mathcal{N}(0, \sigma^2) \) (in our case, \( \sigma = 1 \)). Since \( \mathbb{E} [Z_{(1)}] \leq \sqrt{2 \log M} \), it holds that
\[
\Pr \left\{ Z_{(1)} \geq \sqrt{2 \log M + \xi} \right\} \leq 2e^{-\xi^2}.
\]
Therefore, define \( \mathcal{E}_2 := \{ Z_{(1)} \geq \sqrt{2 \log M + \xi} \} \) and we obtain
\[
\mathbb{E}_{Z_1, \ldots, Z_M \sim \mathcal{N}(0,1)} \left[ \frac{1}{\sqrt{\sum_{i=1}^M Z_i^2}} \sum_{i=1}^k Z_i \right] \leq k \mathbb{E}_{Z_1, \ldots, Z_M \sim \mathcal{N}(0,1)} \left[ \frac{1}{\sqrt{\sum_{i=1}^M Z_i^2}} Z_{(1)} \right] \leq k \cdot \left( \mathbb{E} \left[ \frac{Z_{(1)}}{\sqrt{\sum_{i=1}^M Z_i^2}} \right] \mathcal{E}_1 \mathcal{E}_2 \right) + \sup_{z_1, \ldots, z_m} \left( \frac{z_{(1)}}{\sqrt{\sum_{i=1}^M z_i^2}} \right) \cdot \Pr (\mathcal{E}_1^c \cup \mathcal{E}_2^c) \leq k \cdot \left( \sqrt{2 \log M + \frac{\xi}{M(1-\gamma)}} + 1 \cdot \left( e^{-M\gamma^2/4} + 2e^{-\xi^2} \right) \right) \leq k \cdot \left( \sqrt{2 \log M + \log(M)} \right) + 1 \cdot \left( e^{-M/400} + 2/M \right) = \Theta \left( \frac{k \sqrt{\log M}}{M} \right),
\]
where the last inequality holds by picking \( \gamma = 0.1 \) and \( \xi = \sqrt{\log M} \).

**Lower bound.** The analysis of the lower bound is more sophisticated. To begin with, let
\[
\mathcal{E}_M := \left\{ (Z_1, \ldots, Z_M) \mid \sum_{i=1}^M Z_i^2 \in [M(1-\gamma), M(1+\gamma)] \right\}
\]
denote the good event such that the denominator of our target is well-controlled, where \( \gamma > 0 \) again will be optimized later. By the concentration of \( \chi^2 \) random variables, it holds that
\[
\Pr \{ \mathcal{E}_M^c \} \leq e^{-\frac{M}{2}(\gamma - \log(1+\gamma))} + e^{-M/2} \left( 1 - \frac{1}{\sqrt{\frac{1}{1+\gamma}}} \right) \gamma + e^{-M/2} \leq 2e^{-M/2}. \tag{101}
\]
Next, to lower bound \( \sum_{i=1}^k Z_{(i)} \), we partition \((Z_1, Z_2, \ldots, Z_M)\) into \( k \) blocks \( B_1, B_2, \ldots, B_k \) where each block contains at least \( N = \lceil M/k \rceil \) samples: \( B_j := \{(j-1) \cdot N + 1 : j \cdot N\} \) for \( j \in [k-1] \) and \( B_k = [M] \setminus \bigcup_{j=1}^{k-1} B_j \). Define \( \hat{Z}_{(1)}^{(j)} \) be the maximum samples in the \( j \)-th block: \( \hat{Z}_{(1)}^{(j)} := \max_{i \in B_j} Z_i \). Then, it is obvious that
\[
\sum_{i=1}^k Z_{(i)} \geq \sum_{j=1}^k \hat{Z}_{(1)}^{(j)},
\]
To this end, we define \( \mathcal{E}_1 \) to be the good event that 90\% of \( \hat{Z}_{(1)}^{(j)} \)'s are large enough (i.e., concentrated to the expectation):
\[
\mathcal{E}_1 := \left\{ \left\{ j \in [k] \mid \hat{Z}_{(1)}^{(j)} \geq \frac{\log N}{\sqrt{\pi \log 2}} - \log 100 \right\} \right\} \geq 0.9k.
\]
Note that by the Borell-TIS inequality, for any \( j \in [k] \),
\[
\Pr \left\{ \hat{Z}_{(1)}^{(j)} \geq \frac{\log N}{\sqrt{\pi \log 2}} - \xi \right\} \geq 1 - 2e^{-\xi^2},
\]

18
so setting $\xi = \log 100$ implies $\Pr \left\{ \hat{Z}(j) \geq \frac{\log N}{\sqrt{\pi \log 2}} - \xi \right\} \geq 0.98$. Since blocks are independent with each other, applying Hoeffding’s bound yields

$$\Pr \left\{ \mathcal{E}_1 \right\} \geq 1 - \Pr \left\{ \text{Binom}(k, 0.98) \leq 0.9 \right\} \geq 1 - e^{-k(0.08)^2} \geq 0.9,$$

when $k \geq 400 \cdot \log 10 \geq \log 10/0.08^2$.

Next, we define a “not-too-bad” event where $\sum_{j=1}^{k} \hat{Z}(j)_{(1)}$ is not catastrophically small:

$$\mathcal{E}_2 := \left\{ \sum_{j=1}^{k} \hat{Z}(j)_{(1)} \geq -\frac{k}{\sqrt{M}} \xi \right\},$$

for some $\xi > 0$ to be optimized later. Observe that

$$\Pr \left\{ \mathcal{E}_2 \right\} \geq \Pr \left\{ \frac{k}{M} \sum_{i=1}^{M} Z_i \geq -\frac{k}{\sqrt{M}} \xi \right\} \geq 1 - e^{-\xi^2/2},$$

where (a) holds since the each of the top-$k$ values must be greater than $k$ times the average, and (b) holds due to the Hoeffding’s bound on the sum of i.i.d. Gaussian variables.

Lastly, a trivial bound implies that

$$\inf_{a \in \mathbb{U}} \sum_{i=1}^{k} \alpha(i) \geq -\frac{k}{\sqrt{M}}.$$

Now, we are ready to bound $\mathbb{E}_{z_1, \ldots, z_M \sim \mathcal{N}(0, 1)} \left[ \frac{1}{\sqrt{\sum_{i=1}^{M} z_i^2}} \sum_{i=1}^{k} z_{(i)} \right]$. We begin by decomposing it into three parts:

$$\mathbb{E}_{z_1, \ldots, z_M \sim \mathcal{N}(0, 1)} \left[ \frac{\sum_{i=1}^{k} Z_{(i)}}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \right] = \Pr \left\{ \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_M \right\} \cdot \mathbb{E} \left[ \frac{\sum_{i=1}^{k} Z_{(i)}}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \middle| \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_M \right]$$
$$+ \Pr \left\{ \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_M \right\} \cdot \mathbb{E} \left[ \frac{\sum_{i=1}^{k} Z_{(i)}}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \middle| \mathcal{E}_1^c \cap \mathcal{E}_2 \cap \mathcal{E}_M \right]$$
$$+ \Pr \left\{ \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_M \right\} \cdot \mathbb{E} \left[ \frac{\sum_{i=1}^{k} Z_{(i)}}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \middle| \mathcal{E}_2^c \cup \mathcal{E}_M^c \right].$$

We bound these three terms separately. To bound the first one, observe that condition on $\mathcal{E}_1 \cap \mathcal{E}_2$, $\sum_{i=1}^{k} Z_{(i)} \geq \hat{Z}(j)_{(1)} \geq 0.9k \sqrt{\frac{\log N}{\pi \log 2}} - \frac{k}{\sqrt{M}} \gamma$. As a result,

$$\Pr \left\{ \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_M \right\} \cdot \mathbb{E} \left[ \frac{\sum_{i=1}^{k} Z_{(i)}}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \middle| \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_M \right] \geq \frac{0.9k \sqrt{\frac{\log N}{\pi \log 2}} - \frac{k}{\sqrt{M}} \gamma}{\sqrt{M(1 + \gamma)}} \cdot \left( 1 - \left( 0.1 + e^{-\xi^2/2} + 2e^{-M\gamma^2/4} \right) \right).$$

To bound the second term, observe that under $\mathcal{E}_2$,

$$\sum_{i=1}^{k} Z_{(i)} \geq -\frac{k}{\sqrt{M}} \xi,$$
so we have
\[
\Pr \{ E_2 \cap E_1 \cap E_M \} \cdot \mathbb{E} \left[ \frac{\sum_{i=1}^{k} Z_{(i)}^{2}}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \middle| E_2 \cap E_1 \cap E_M \right] \\
\geq \Pr \{ E_2 \cap E_1 \cap E_M \} \cdot \left( -\frac{k}{\sqrt{M^2(1-\gamma)}} \xi \right) \\
\geq \Pr \{ E_1 \} \cdot \left( -\frac{k}{\sqrt{M^2(1-\gamma)}} \xi \right) \\
\geq 0.1 \cdot \left( -\frac{\xi \sqrt{k}}{\sqrt{M^2(1-\gamma)}} \right) .
\]  
(103)

For the third term, it holds that
\[
\Pr \{ E_2 \cup E_M \} \cdot \mathbb{E} \left[ \frac{\sum_{i=1}^{k} Z_{(i)}^{2}}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \middle| E_2 \cup E_M \right] \\
\geq \Pr \{ E_2 \cup E_M \} \cdot \inf_{a \in \mathbb{S}^{M-1}} \sum_{i=1}^{k} a_{(i)} \\
\geq -\Pr \{ E_2 \cup E_M \} \cdot \frac{k}{\sqrt{M}} \\
\geq - \left( e^{-\xi^2/2} + e^{-M\gamma^2/4} \right) \cdot \frac{k}{\sqrt{M}}.
\]  
(104)

Combining (102), (103), and (104) together, we arrive at
\[
\mathbb{E}_{Z_1,\ldots,Z_M \sim N(0,1)} \left[ \frac{\sum_{i=1}^{k} Z_{(i)}^{2}}{\sqrt{\sum_{i=1}^{M} Z_i^2}} \right] \\
\geq 0.9k \left( \frac{\log N}{\pi \log 2} \right) \frac{k}{\sqrt{M(1+\gamma)}} \gamma \left( 1 - \left( 0.1 + e^{-\xi^2/2} + e^{-M\gamma^2/4} \right) \right) \\
- \left( e^{-\xi^2/2} + e^{-M\gamma^2/4} \right) \cdot \frac{k}{\sqrt{M}}.
\]

Finally, setting \( \gamma = O \left( \frac{1}{\sqrt{M}} \right) \) and \( \xi = O(1) \) yields the desired lower bound
\[
C_{d,M,k} = \Omega \left( \frac{k \log N}{\sqrt{M}} \right).
\]

\( \square \)

H Additional Experimental Results

In Figure 2 we provide additional empirical results by sweeping the number of users \( n \) from 2,000 to 10,000 on the left and sweeping the dimension \( d \) from 200 to 1,000 on the right.

I Additional Details on Prior LDP Schemes

For completeness, we provide additional details on prior LDP mean estimation schemes in this section, including PrivUnit \([4]\), SQKR \([6]\), FT21 \([12]\), and MMRC \([30]\). We skip prior work analyzing compression-privacy-utility tradeoffs that do not specifically focus on the distributed mean estimation problem \([19, 20]\) or others that study frequency estimation \([6, 11, 30]\).
The authors showed that PrivUnit is an exact optimal among the family of unbiased locally private

$$N, R$$ represent the local vector $$x$$ that if $$P$$ Parseval’s identity, i.e. $$\sum_{j=1}^{N} (u_j, v)^2$$ for all $$v \in \mathbb{R}^d$$. We say that the expansion $$v = \sum_{j=1}^{N} a_j u_j$$ is a Kashin’s representation of $$x$$ at level $$K$$ if max $$|a_j| \leq \frac{K}{\sqrt{N}} \|v\|_2$$ [23]. [27] shows that if $$N > (1 + \mu) d$$ for some $$\mu > 0$$, then there exists a tight frame $$\{u_j\}_{j=1}^{N}$$ such that for any $$x \in \mathbb{R}^d$$, one can find a Kashin’s representation at level $$K = \Theta(1)$$. This implies that we can represent the local vector $$v$$ with coefficients $$\{a_j\}_{j=1}^{N} \in [-c/\sqrt{d}, c/\sqrt{d}]^N$$ for some constants $$c$$ and $$N = \Theta(d)$$. 

I.1 PrivUnit [4]

[2] considered the mean estimation problem under DP constraint (without communication constraint) when $$X = \mathbb{S}^{d-1} = \{v \in \mathbb{R}^d : \|v\|_1 = 1\}$$. Since there is no communication constraint, they assumed canonical protocol where the random encoder is $$\mathbb{S}^{d-1}$$ considered the mean estimation problem under DP constraint (without communication constraint) $$Y = \mathbb{S}^{d-1} = \{v \in \mathbb{R}^d : \|v\|_1 = 1\}$$, with $$S_{-1}$$ denoting the surface area of hypersphere cap $$\{z \in \mathbb{S}^{d-1} | \langle z, v \rangle \geq \gamma\}$$, with $$S_{-1}$$ representing the surface area of the $$d$$ dimensional hypersphere. We denote $$q = \Pr [Z_1 \leq \gamma] = (S_{-1} - S_1)/S_{-1}$$ (convention from [3] [2]). The normalization factor is required to obtain unbiasedness.

[2] also introduced PrivUnitG, which is a Gaussian approximation of PrivUnit. In this approach, $$Z$$ is sampled from an i.i.d. $$\mathcal{N}(0, 1/d)$$ distribution. This simplifies the process of determining more accurate parameters $$p, q, \gamma$$. Consequently, in practical applications, PrivUnitG surpasses PrivUnit in performance owing to superior parameter optimization.

I.2 SQKR [6]

Next, we outline the encoder and decoder of SQKR in this section. The encoding function mainly consists of three steps: (1) computing Kashin’s representation, (2) quantization, and (3) sampling and privatization.

**Compute Kashin’s representation** A tight frame is a set of vectors $$\{u_j\}_{j=1}^{N} \in \mathbb{R}^d$$ that satisfy Parseval’s identity, i.e. $$\|v\|_2^2 = \sum_{j=1}^{N} (u_j, v)^2$$ for all $$v \in \mathbb{R}^d$$. We say that the expansion $$v = \sum_{j=1}^{N} a_j u_j$$ is a Kashin’s representation of $$x$$ at level $$K$$ if max $$|a_j| \leq \frac{K}{\sqrt{N}} \|v\|_2$$ [23]. [27] shows that if $$N > (1 + \mu) d$$ for some $$\mu > 0$$, then there exists a tight frame $$\{u_j\}_{j=1}^{N}$$ such that for any $$x \in \mathbb{R}^d$$, one can find a Kashin’s representation at level $$K = \Theta(1)$$. This implies that we can represent the local vector $$v$$ with coefficients $$\{a_j\}_{j=1}^{N} \in [-c/\sqrt{d}, c/\sqrt{d}]^N$$ for some constants $$c$$ and $$N = \Theta(d)$. 

Figure 2: Comparison of RRSC with SQKR [6], MMRC [30], and PrivUnitG [2]. (left) $$\ell_2$$ error vs number of users $$n$$ with $$d = 500, \varepsilon = 6$$, and the number of bits is $$b = \varepsilon = 6, k = 1$$ for each $$n$$. (right) $$\ell_2$$ error vs dimension $$d$$ for $$n = 5000, \varepsilon = 6$$, and the number of bits is $$b = \varepsilon = 6, k = 1$$ for each $$d$$. 

![Comparison of RRSC with SQKR, MMRC, and PrivUnitG](image)
Quantization. In the quantization step, each client quantizes each $a_j$ into a 1-bit message $q_j \in \{-c/\sqrt{d}, c/\sqrt{d}\}$ with $\mathbb{E}[q_j] = a_j$. This yields an unbiased estimator of $\{a_j\}_{j=1}^N$, which can be described in $N = \Theta(d)$ bits. Moreover, due to the small range of each $a_j$, the variance of $q_j$ is bounded by $O(1/d)$.

Sampling and privatization. To further reduce $\{q_j\}$ to $k = \min(\lfloor \varepsilon \rfloor, b)$ bits, client $i$ draws $k$ independent samples from $\{q_j\}_{j=1}^N$ with the help of shared randomness, and privatizes its $k$ bits message via $2^k$-RR mechanism \cite{tcc19}, yielding the final privatized report of $k$ bits, which it sends to the server.

Upon receiving the report from client $i$, the server can construct unbiased estimators $\hat{a}_j$ for each $\{a_j\}_{j=1}^N$, and hence reconstruct $\hat{v} = \sum_{j=1}^N \hat{a}_j u_j$, which yields an unbiased estimator of $v$. In \cite{tcc19}, it is shown that the variance of $\hat{v}$ can be controlled by $O(d/\min(\varepsilon^2, \varepsilon, b))$.

I.3 FT21 \cite{tcc19} and MMRC \cite{tcc19}

Both FT21 and MMRC aim to simulate a given $\varepsilon$-LDP scheme. More concretely, consider an $\varepsilon$-LDP mechanism $q(\cdot|v)$ that we wish to compress, which in our case, $\text{PrivUnit}$. A number of candidates $u_1, \cdots, u_N$ are drawn from a fixed reference distribution $p(u)$ (known to both the client and the server), which in our case, uniform distribution on the sphere $S^{d-1}$. Under FT21 \cite{tcc19}, these candidates are generated from an (exponentially strong) PRG, with seed length $\ell = \text{polylog}(d)$. The client then performs rejection sampling and sends the seed of the sampled candidates to the server. See Algorithm 2 for an illustration.

Algorithm 2 Simulating LDP mechanisms via rejection sampling \cite{tcc19}

Inputs: $\varepsilon$-LDP mechanism $q(\cdot|v)$, ref. distribution $p(\cdot)$, seeded PRG $G : \{0,1\}^\ell \rightarrow \{0,1\}^\ell$, failure probability $\gamma > 0$, $J = e^\varepsilon \ln(1/\gamma)$.

for $j \in \{1, \cdots, J\}$ do

Sample a random seed $s \in \{0,1\}^\ell$.

Draw $u \leftarrow p(\cdot)$ using the PRG $G$ and the random seed $s$.

Sample $b$ from Bernoulli $\left(\frac{q(u|v)}{e^\varepsilon p(u)}\right)$.

if $b = 1$ then

BREAK

end if

end for

Output: $s$

On the other hand, under MMRC \cite{tcc19} the LDP mechanism is simulated via a minimal random coding technique \cite{tcc19}. Specifically, the candidates are generated via shared randomness, and the client performs an importance sampling and sends the index of the sampled one to the server, as illustrated in Algorithm 3. It can be shown that when the target mechanism is $\varepsilon$-LDP, the communication costs of both strategies are $\Theta(\varepsilon)$ bits. It is also worth noting that both strategies will incur some bias (though the bias can be made exponentially small as one increases the communication cost), and \cite{tcc19} provides a way to correct the bias when the target mechanism is PrivUnit (or general cap-based mechanisms).
Algorithm 3 Simulating LDP mechanisms via importance sampling \[30\]

**Inputs:** \(\varepsilon\)-LDP mechanism \(q(\cdot|v)\), ref. distribution \(p(\cdot)\), # of candidates \(M\)

1. Draw samples \(u_1, \cdots, u_M\) from \(p(u)\) using the shared source of randomness.
2. For \(k \in \{1, \cdots, M\}\) do
   - \(w(k) \leftarrow q(u_k|v)/p(u_k)\).
3. Let \(\pi_{\text{HBC}}(\cdot) \leftarrow w(\cdot)/\sum_k w(k)\).
   - Draw \(k^* \leftarrow \pi_{\text{HBC}}\).

**Output:** \(k^*\)
References


