Appendix

Table of Contents

A Fully interactive model

In this appendix, we describe how to extend our results, presented in the sequentially interactive model, to the more general interactive setting. We first formally define this setting and the corresponding notion of protocols. Hereafter, we use * for the Kleene star operation, *i.e.*, $V^* = \bigcup_{n=0}^{\infty} V^n$.

Definition 4 (Interactive Protocols). Let $X_1, \dots, X_n$ be i.i.d. samples from $\mathbf{p}_\theta$, $\theta \in \Theta$, and $\mathcal{W}^*$ be a collection of sequences of pairs of channel families and players; that is, each element of $\mathcal{W}^*$ is a sequence $(\mathcal{W}_t, j_t)_{t\in\mathbb{N}}$ where $j_t \in [n]$. An *interactive protocol* Π *using* $\mathcal{W}^*$ comprises a random variable U (independent of the input $X_1, \dots, X_n$) and, for each $t \in \mathbb{N}$, mappings

$$\sigma_t \colon Y_1, \dots, Y_{t-1}, U \mapsto N_t \in [n] \cup \{\bot\}$$
$$g_t \colon Y_1, \dots, Y_{t-1}, U \mapsto W_t$$

with the constraint that $((W_1, N_1), \dots, (W_t, N_t))$ must be consistent with some sequence from $\mathcal{W}^*$; that is, there exists $((\mathcal{W}_s, j_s))_{s\in\mathbb{N}} \in \mathcal{W}^*$ such that $W_s \in \mathcal{W}_s$ and $N_s = j_s$ for all $1 \le s \le t$. These two mappings respectively indicate (i) whether the protocol is to stop (symbol $\bot$), and, if not, which player is to speak at round $t \in \mathbb{N}$, and (ii)) which channel this player selects at this round.

In round t, if $N_t = \bot$, the protocol ends. Otherwise, player N_t (as selected by the protocol, based on the previous messages) uses the channel W_t to produce the message (output) Y_t according to the probability measure $W_t(\cdot \mid X_{N_t})$. We further require that $T := \inf\{\, t \in \mathbb{N} : \ N_t = \bot \,\}$ is finite a.s. The messages $Y^T = (Y_1, \dots, Y_T)$ received by the referee and the public randomness U constitute the *transcript* of the protocol Π.

In other terms, the channel used by the player N_t speaking at time t is a Markov kernel

$$W_t\colon \mathfrak{Y}_t \times \mathcal{X} \times \mathcal{Y}^{t-1} \to [0,1]\,,$$

with $\mathcal{Y}_t \subseteq \mathcal{Y}$; and, for player $j \in [n]$, the allowed subsequences $(\mathcal{W}_t, j_t)_{t\in\mathbb{N}: j_t=j}$ capture the possible sequences of channels allowed to the player. As an example, if we were to require that any single player can speak at most once, then for every $j \in [n]$ and every $(\mathcal{W}_t, j_t)_{t\in\mathbb{N}} \in \mathcal{W}^n$, we would have $\sum_{t=1}^{\infty} \mathbb{1}\{j_t = j\} \le 1$.

In the interactive model, we can then capture the constraint that each player must communicate at most ℓ bits in total by letting $\mathcal{W}^n$ be the set of sequences $(\mathcal{W}_t^{\text{comm},\ell_t}, j_t)_{t\in\mathbb{N}}$ such that

$$\forall j \in [n], \qquad \sum_{t=1}^{\infty} \ell_t \cdot \mathbb{1}\{j_t = j\} \le \ell\,.$$

In the simpler sequentially interactive model, this condition simply becomes the choice of $\mathcal{W}^n = (\mathcal{W}^{\text{comm},\ell}, \dots, \mathcal{W}^{\text{comm},\ell})$.

A.1 Lower Bounds under Full Interactive Model

Next we discuss how our technique extends to the full interactive model. For any full interactive protocol Π, let $Y^* \in \mathcal{Y}^*$ be the message sequence generated by the protocol. Then, for all $y^* \in \mathcal{Y}^*$, we have

$$\Pr_{X^n\sim\mathbf{p}}[\,Y^* = y^*\,] = \mathbb{E}_{X^n\sim\mathbf{p}}\left[\prod_{t=1}^{\infty} W_t\big(y_t \mid X_{\sigma_t(y^{t-1})}, y^{t-1}\big)\right].$$

The following lemma states that if X^n are generated from a product distribution, the distribution of the transcript satisfies a property similar to the "cut-and-paste" property from [6].

Lemma 2 ([20])**.** *If $X^n \sim \mathbf{p} = \otimes_{t=1}^n \mathbf{p}_t$, the transcript of the protocol satisfies*

$$\Pr_{X^n\sim\mathbf{p}}[\,Y^* = y^*\,] = \prod_{t=1}^{n} \mathbb{E}_{X_t\sim\mathbf{p}_t}[g_t(y^*, X_t)], \tag{16}$$

where $g_t(y^*, x_t) = \prod_{j=1}^{\infty} W_j(y_j \mid x_t, y^{j-1})\mathbb{1}\{\sigma_j(y^{j-1}) = t\}$.

Hence, when $X^n \sim \mathbf{p}_z^{\otimes n}$ we have

$$\mathbf{p}_z^{y^*} := \Pr_{X^n\sim\mathbf{p}_z^{\otimes n}}[\,Y^* = y^*\,] = \prod_{t=1}^{n} \mathbb{E}_{X_t\sim\mathbf{p}_z}[g_t(y^*, X_t)].$$

Here we can define a similar notion of "channel" for a communication protocol Π for the ith player when the underlying distribution is $\mathbf{p}_z$ by setting

$$\tilde{W}_{t,\mathbf{p}_z}(y^* \mid x) = g_t(y^*, x)\left(\prod_{j\ne t} \mathbb{E}_{X_j\sim\mathbf{p}_z}[g_j(y^*, X_j)]\right). \tag{17}$$

Then we have, for all $t \in [n]$,

$$\mathbb{E}_{X_t\sim\mathbf{p}_z}\left[\tilde{W}_{t,\mathbf{p}_z}(y^* \mid X_t)\right] = \Pr_{X^n\sim\mathbf{p}_z^{\otimes n}}[\,Y^* = y^*\,].$$

We proceed to prove a bound similar to Theorem 1 in terms of the "channel" defined in Eq. (17), as stated below.

Theorem 4 (Information contraction bound)**.** *Fix $\tau \in (0, 1/2]$. Let Π be a fully interactive protocol using $\mathcal{W}^n$, and let Z be a random variable on $\mathcal{Z}$ with distribution $\operatorname{Rad}(\tau)^{\otimes k}$. Let (Y^*, U) be the transcript of Π when the input $X_1, \dots, X_n$ is i.i.d. with common distribution $\mathbf{p}_Z$. Then, under Assumption 1,*

$$\left(\frac{1}{k}\sum_{i=1}^{k} \mathrm{d_{TV}}\Big(\mathbf{p}_{+i}^{Y^*}, \mathbf{p}_{-i}^{Y^*}\Big)\right)^2 \le \frac{7}{k}\alpha^2 \sum_{j=1}^{n} \max_{z\in\mathcal{Z}} \max_{(\mathcal{W}_t, j_t)_{t\in\mathbb{N}}\in\mathcal{W}^n} \sum_{i=1}^{k} \int_{y^*\in\mathcal{Y}^*} \frac{\mathbb{E}_{\mathbf{p}_z}\left[\phi_{z,i}(X)\tilde{W}_{j,\mathbf{p}_z}(y^* \mid X)\right]^2}{\mathbb{E}_{\mathbf{p}_z}\left[\tilde{W}_{j,\mathbf{p}_z}(y^* \mid X)\right]}\, \mathrm{d}\mu\ ,$$

where $\mathbf{p}_{+i}^{Y^*} := \mathbb{E}\big[\mathbf{p}_Z^{Y^*} \mid Z_i = 1\big]$, $\mathbf{p}_{-i}^{Y^*} := \mathbb{E}\big[\mathbf{p}_Z^{Y^*} \mid Z_i = 1\big]$.

We can see the bound is in identical form to Theorem 1 except that we replace each player's channel with the $\tilde{W}_{j,\mathbf{p}_z}(y^* \mid X)$ we defined. Other similar bounds in Section 3 can also be derived under additional assumptions and specific constraints. We present the proof for Theorem 4 below and omit the detailed statements and proof for other bounds.

Proof. Analogously to Eq. (33), we can get

$$\frac{1}{k}\left(\sum_{i=1}^{k} \mathrm{d}_{\mathrm{TV}}\left(\mathbf{p}_{+i}^{Y^*}, \mathbf{p}_{-i}^{Y^*}\right)\right)^2 \le 14 \sum_{t=1}^{n} \mathbb{E}_Z\left[\sum_{i=1}^{k} \mathrm{d}_{\mathrm{H}}\left(\mathbf{p}_Z^{Y^*}, \mathbf{p}_{t\leftarrow Z\oplus i}^{Y^*}\right)^2\right] \tag{18}$$

For all $z \in \{-1,+1\}^k$ and i, t, by the definition of Hellinger distance and Eq. (16), we have

$$\begin{aligned} 2\mathrm{d}_{\mathrm{H}}\left(\mathbf{p}_z^{Y^*}, \mathbf{p}_{t\leftarrow z\oplus i}^{Y^*}\right)^2 &= \int_{y^*\in\mathcal{Y}^*} \prod_{\substack{1\le j\le n\\ j\ne t}} \mathbb{E}_{X_j\sim\mathbf{p}_z}[g_j(y^*, X_j)] \left(\sqrt{\mathbb{E}_{X_t\sim\mathbf{p}_{z\oplus i}}[g_t(y^*, X_t)]} - \sqrt{\mathbb{E}_{X_t\sim\mathbf{p}_z}[g_t(y^*, X_t)]}\right)^2 \mathrm{d}\mu \\ &\le \int_{y^*\in\mathcal{Y}^*} \left(\prod_{j\ne t} \mathbb{E}_{X_j\sim\mathbf{p}_z}[g_j(y^*, X_j)]\right)\left(\frac{(\mathbb{E}_{X_t\sim\mathbf{p}_z}[g_t(y^*, X_t)] - \mathbb{E}_{X_t\sim\mathbf{p}_{z\oplus i}}[g_t(y^*, X_t)])^2}{\mathbb{E}_{X_t\sim\mathbf{p}_z}[g_t(y^*, X_t)]}\right)\mathrm{d}\mu\,, \end{aligned}$$

Proceeding from above, we get under Assumption 1,

$$\begin{aligned} 2\mathrm{d}_{\mathrm{H}}\left(\mathbf{p}_z^{Y^*}, \mathbf{p}_{t\leftarrow z\oplus i}^{Y^*}\right)^2 &\le \alpha^2 \int_{y^*\in\mathcal{Y}^*} \left(\prod_{j\ne t} \mathbb{E}_{X_j\sim\mathbf{p}_z}[g_j(y^*, X_j)]\right)\left(\frac{\mathbb{E}_{X_t\sim\mathbf{p}_z}[\phi_{z,i}(X_t) g_t(y^*, X_t)]^2}{\mathbb{E}_{X_t\sim\mathbf{p}_z}[g_t(y^*, X_t)]}\right)\mathrm{d}\mu \\ &= \alpha^2 \int_{y^*\in\mathcal{Y}^*} \frac{\mathbb{E}_{X_t\sim\mathbf{p}_z}\left[\phi_{z,i}(X_t) g_t(y^*, X_t)\prod_{j\ne t}\mathbb{E}_{X_j\sim\mathbf{p}_z}[g_j(y^*, X_j)]\right]^2}{\mathbb{E}_{X_t\sim\mathbf{p}_z}\left[g_t(y^*, X_t)\prod_{j\ne t}\mathbb{E}_{X_j\sim\mathbf{p}_z}[g_j(y^*, X_j)]\right]}\mathrm{d}\mu \\ &= \alpha^2 \int_{y^*\in\mathcal{Y}^*} \frac{\mathbb{E}_{X_t\sim\mathbf{p}_z}\left[\phi_{z,i}(X_t)\tilde{W}_{t,\mathbf{p}_z}(y^* \mid X)\right]^2}{\mathbb{E}_{X_t\sim\mathbf{p}_z}\left[\tilde{W}_{t,\mathbf{p}_z}(y^* \mid X)\right]}\mathrm{d}\mu\,. \end{aligned}$$

Plugging the above bound into Eq. (18), we can obtain the bound in Theorem 4 by taking the maximum over all $z \in \{-1,+1\}^k$ and all possible channel sequences. □

B A measure change bound

We here provide a variant of Talagrand's transportation-cost inequality which is used in deriving Eq. (5) (under Assumption 3) in the second part of Theorem 2. We note that this type of result is not novel, and can be derived from standard arguments in the literature (see, e.g., [9, Chapter 8] or [27, Chapter 4]). However, the lemma below is specifically tailored for our purposes, and we provide the proof for completeness. A similar bound was derived in [2], where Gaussian mean testing under communication constraints was considered.

Lemma 3 (A measure change bound)**.** *Consider a random variable X taking values in $\mathcal{X}$ and with distribution P. Let $\Phi\colon \mathcal{X} \to \mathbb{R}^k$ be such that the random vector $\Phi(X)$ is σ^2-subgaussian. Then, for any function $a\colon \mathcal{X} \to [0,\infty)$ such that $\mathbb{E}[a(X)] < \infty$, we have*

$$\frac{\|\mathbb{E}[\Phi(X)a(X)]\|_2^2}{\mathbb{E}[a(X)]^2} \le 2\sigma^2 \frac{\mathbb{E}[a(X)\ln a(X)]}{\mathbb{E}[a(X)]} + 2\sigma^2 \ln\frac{1}{\mathbb{E}[a(X)]}.$$

Proof. By an application of Gibb's variational principle ($cf.$ [9, Corollary 4.14]) the following holds: For a random variable Z and distributions P and Q on the underlying probability space satisfying $Q \ll P$ (that is, such that Q is absolutely continuous with respect to P), we have

$$\lambda\mathbb{E}_Q[Z] \le \ln\mathbb{E}_P\left[e^{\lambda Z}\right] + \mathrm{D}(Q\|P).$$

To apply this bound, set P to be the distribution of X and let $Q \ll P$ be defined using its density (Radon–Nikodym derivative) with respect to P given by

$$\frac{\mathrm{d}Q}{\mathrm{d}P} = \frac{a(X)}{\mathbb{E}_P[a(X)]}.$$

Now, note that for any unit vector v, we have, setting $Z = v^\intercal \Phi(X)$ and using the σ^2-subgaussianity of $\Phi(X)$, that

$$\lambda \mathbb{E}_Q[v^\intercal \Phi(X)] \leq \ln \mathbb{E}_P\left[e^{\lambda v^\intercal \Phi(X)}\right] + \mathrm{D}(Q\|P) \leq \frac{\sigma^2\lambda^2}{2} + \mathrm{D}(Q\|P).$$

In particular, for $\lambda = \frac{1}{\sigma}\sqrt{2\mathrm{D}(Q\|P)}$, we get

$$\mathbb{E}_Q[v^\intercal \Phi(X)] \leq \sigma\sqrt{2\mathrm{D}(Q\|P)}.$$

Applying this to the unit vector $v := \frac{\mathbb{E}_Q[\Phi(X)]}{\|\mathbb{E}_Q[\Phi(X)]\|_2}$ then yields

$$\|\mathbb{E}_Q[\Phi(X)]\|_2 \leq \sigma\sqrt{2\mathrm{D}(Q\|P)}.$$

To conclude, it then suffices to observe that

$$\mathrm{D}(Q\|P) = \frac{\mathbb{E}_P[a(X)\ln a(X)]}{\mathbb{E}_P[a(X)]} + \ln\frac{1}{\mathbb{E}_P[a(X)]}.$$

The proof is completed by combining the bounds above, as $\mathbb{E}_Q[\Phi(X)] = \frac{\mathbb{E}_P[\Phi(X)a(X))]}{\mathbb{E}_P[a(X)]}$. □

C Upper bounds

We now describe and analyze the interactive algorithms for the estimation tasks we consider.

C.1 Product Bernoulli Distributions

Recall that $\mathcal{B}_{d,s}$, the family of d-dimensional s-sparse product Bernoulli distributions, is defined as

$$\mathcal{B}_{d,s} := \left\{ \bigotimes_{j=1}^{d} \mathrm{Rad}(\frac{1}{2}(\mu_j + 1)) \ : \ \mu \in [-1,1]^d, \|\mu\|_0 \leq s \right\}. \tag{19}$$

We now provide the interactive protocols achieving the upper bounds of Theorem 3 for sparse product Bernoulli mean estimation under LDP and communication constraints .

Our protocols has two ingredients described below:

- **Estimating non-zero mean coordinates.**In this step we will start with $S_0 = [d]$, the set of all possible coordinates. Then we will iteratively prune the set $S_0 \to S_1 \to \ldots \to S_T$, such that $|S_T| = 3s$ (this step is skipped if $s \geq d/3$) is a good estimate for the set of coordinates with non-zero mean.
- **Estimating the non-zero means.** We then estimate the means of the coordinates in S_T, which is equivalent to solving a dense mean estimation problem in $3s$ dimensions.

In the next two sections, we provide the details of the algorithm that matches the lower bounds obtained in Section 5 for interactive protocols under LDP and communication constraints respectively.

C.1.1 LDP constraints

In this subsection, we will focus on the case $\varepsilon \in (0, 1]$ (high-privacy regime). For the case $\varepsilon > 1$, we rely a privatization of the communication-limited algorithm, which will be discussed at the end of Appendix C.1.2. Our protocol for Bernoulli mean estimation under LDP constraints is described in Algorithm 1. As stated above, in each round $t = 1, \ldots, T$, for each $j \in S_{t-1}$ a new group of players apply the well known binary Randomized Response (RR) mechanism [29, 24] to their jth coordinate. Using these messages we then guess a set of coordinates with highest possible means (in absolute value) and prune the set to S_t. This is done in Lines 2-6 of Algorithm 1.

In Lines 7-12, the algorithm uses the same approach to estimate the means of coordinates within S_T and sets remaining coordinates to zero.

The privacy guarantee follows immediately from that of the RR mechanism, and further, this only requires one bit of communication per player.

Algorithm 1 LDP protocol for mean estimation for the product of Bernoulli family

Require: n players, dimension d, sparsity parameter s, privacy parameter ε.

1: Set $T := \log_3 \frac{d}{3s}$, $\alpha := \frac{e^\varepsilon}{1+e^\varepsilon}$, $S_0 = [d]$, $N_0 := \frac{n}{6d}$.
2: **for** $t = 1, 2, \ldots, T$ **do**
3: **for** $j \in S_{t-1}$ **do**
4: Get a group of new players $G_{t,j}$ of size $N_t = N_0 \cdot 2^t$.
5: Player $i \in G_{t,j}$, upon observing $X_i \in \{-1,+1\}^d$ sends the message $Y_i \in \{-1,+1\}$ such that

$$Y_i = \begin{cases} (X_i)_j & \text{w.p. } \alpha, \\ -(X_i)_j & \text{w.p. } 1-\alpha. \end{cases} \tag{20}$$

6: Set $M_{t,j} := \sum_{i \in G_{t,j}} Y_i$. Let $S_t \subseteq S_{t-1}$ be the set of the $|S_{t-1}|/3$ indices with the largest $|M_{t,j}|$.
7: **for** $j \in S_T$ **do**
8: Get a group of new players $G_{T,j}, j \in S_T$ of size $N_{T+1} = N_0 \cdot 2^T$.
9: Player $i \in G_{T,j}$, sends the message $Y_i \in \{-1,+1\}$ according to Eq. (20) and $M_{T,j} := \sum_{i \in G_{T,j}} Y_i$
10: **for** $j \in [d]$ **do**
11:

$$\widehat{\mu}_j = \begin{cases} \frac{M_{j,T}}{(2\alpha-1)N_{T+1}} & \text{if } j \in S_T, \\ 0 & \text{otherwise.} \end{cases}$$

12: **return** $\widehat{\mu}$.

The performance guarantee of Algorithm 1 is stated below, which matches the lower bounds obtained in Section 5.

Proposition 1. *Fix $p \in [1,\infty]$. For $n \geq 1$ and $\varepsilon \in (0,1]$, Algorithm 1 is an (n,γ)-estimator using $\mathcal{W}_\varepsilon$ under ℓ_p loss for $\mathcal{B}_{d,s}$ with $\gamma = O\left(\sqrt{\frac{pds^{2/p}}{n\varepsilon^2}}\right)$ for $p \leq 2\log s$ and $\gamma = O\left(\sqrt{\frac{d\log s}{n\varepsilon^2}}\right)$ for $p > 2\log s$.*

Proof. The total number of players used by Algorithm 1 uses is

$$\sum_{t=1}^{T+1} |S_{t-1}| \cdot N_t = |S_0| \cdot N_0 \cdot \sum_{t=1}^{T+1} \frac{2^t}{3^{t-1}} \leq 6|S_0| \cdot N_0 = n.$$

To prove the utility guarantee, we bound the estimation error in the estimated set S_T and the error outside the set S_T in the following lemma.

Lemma 4. *Let S_T be the subset obtained from the first stage of Algorithm 1. Then,*

$$\max\left\{\mathbb{E}\left[\sum_{j \notin S_T} |\mu_j - \widehat{\mu}_j|^p\right], \mathbb{E}\left[\sum_{j \in S_T} |\mu_j - \widehat{\mu}_j|^p\right]\right\} = O\left(s\left(\frac{pd}{n\varepsilon^2}\right)^{p/2}\right).$$

The proposition follows directly from the lemma. Indeed, for $p > 2\log s$, by monotonicity of ℓ_p norms we have $\|\mu - \hat{\mu}\|_p \leq \|\mu - \hat{\mu}\|_{p'}$ for all $p' \leq p$, and thus choosing $p' := 2\log s$ is sufficient to obtain the stated bound. □

Proof of Lemma 4. We prove the bound on each term individually. The first term captures the performance of our estimator within coordinates in S_T and the second term states that we do not "prune" too many coordinates with high non-zero means.

Bounding the first term. For $j \notin S_T$, we output $\widehat{\mu}_j = 0$. Therefore,

$$\mathbb{E}\left[\sum_{j \notin S_T} |\mu_j - \widehat{\mu}_j|^p\right] = \sum_j \mathbb{E}[|\mu_j - \widehat{\mu}_j|^p \cdot \mathbb{1}\{j \notin S_T\}] = \sum_j |\mu_j|^p \cdot \Pr[j \notin S_T].$$

Since μ is s-sparse, it will suffice to show that for all j with $|\mu_j| > 0$,

$$|\mu_j|^p \cdot \Pr[\, j \notin S_T\,] = O\left(\left(\frac{pd}{n\varepsilon^2}\right)^{p/2}\right). \tag{21}$$

Let

$$H := 20\sqrt{\frac{d}{n(2\alpha - 1)^2}}.$$

Note that for $\varepsilon \in (0, 1]$, we have $2\alpha - 1 \geq \frac{e-1}{e+1}\varepsilon$. Therefore, if $|\mu_j| \leq H$, then Eq. (21) holds since $\Pr[\, j \notin S\,] \leq 1$. We hereafter assume $|\mu_j| > H$, and let $\mu_j = \beta_j H$ with $\beta_j > 1$. Let $E_{t,j}$ be the event that coordinate j is removed in round t given that $j \in S_{t-1}$. Then we have

$$\Pr[\, j \notin S_T\,] \leq \sum_{t=1}^{T} \Pr[\, E_{t,j}\,].$$

We proceed to bound each $\Pr[\, E_{t,j}\,]$ separately. Note that for $i \in G_{t,j}$, $Y_i \in \{-1, +1\}$ and by Eq. (20)

$$\mathbb{E}[Y_i] = (2\alpha - 1) \cdot \mu_j = (2\alpha - 1)\beta_j H. \tag{22}$$

Let $a_{t,j}$ be the number of coordinates j' with $\mu_{j'} = 0$ and $|M_{t,j'}| \geq \frac{1}{2}N_t(2\alpha - 1)\beta_j H$. Since we select the $|S_{t-1}|/3$ coordinates with the largest magnitude of the sum, for $j \notin S_t$ to happen at least one of the following must occur: (i) $a_{t,j} > \frac{1}{3}|S_{t-1}| - s$, or (ii) $M_{t,j} < \frac{1}{2}N_t(2\alpha - 1)\beta_j H$.

By Hoeffding's inequality, we have

$$\Pr\left[\, M_{t,j} < \frac{1}{2}N_t(2\alpha - 1)\beta_j H\right] \leq \exp\left(-\frac{1}{8}N_t((2\alpha - 1)\beta_j H)^2\right) < \exp\left(-5 \cdot 2^t \beta_j^2\right).$$

Let $p_{t,j} := e^{-5\cdot 2^t \beta_j^2}$. Similarly, for any j' such that $\mu_{j'} = 0$,

$$\Pr\left[\, |M_{t,j'}| \geq \frac{1}{2}N_t(2\alpha - 1)\beta_j H\right] \leq 2p_{t,j}.$$

Since all coordinates are independent, $a_{t,j}$ is binomially distributed with mean at most $2p_{t,j}|S_{t-1}|$. By Markov's inequality, we get

$$\Pr\left[\, a_{t,j} > \frac{1}{3}|S_{t-1}| - s\right] \leq \frac{\mathbb{E}[a_{t,j}]}{|S_{t-1}|/3 - s} \leq p_{t,j},$$

recalling that $|S_{t-1}| = d3^{t-1} \geq 9s$. By a union bound and summing over $t \in [T]$, we get

$$\Pr[\, j \notin S_T\,] \leq \sum_{t=1}^{T} \Pr[\, E_{t,j}\,] \leq \sum_{t=1}^{T} 3p_{t,j} = 3\sum_{t=1}^{T} \exp\left(-2^t \cdot 5\beta_j^2\right) \leq 6\exp\left(-5\beta_j^2\right).$$

Not that for $x > 0$, $x^p e^{-x^2} \leq \left(\frac{p}{2e}\right)^{p/2}$. Hence

$$|\mu_j|^p \cdot \Pr[\, j \notin S_T\,] \leq 6H^p \beta_j^p e^{-5\beta_j^2} \leq \left(C\frac{pd}{n\varepsilon^2}\right)^{p/2},$$

for some absolute constant $C > 0$, completing the proof.

Bounding the second term. Note that S_T is a random variable itself. We show that the bound holds for any realization of S_T. We need the following result which follows from standard moment bounds on binomial distributions.

Fact 1. *Let $p \geq 1$, $m \in \mathbb{N}$, $0 \leq q \leq 1$, and $N \sim \mathrm{Bin}(m, q)$. Then,* $\mathbb{E}[|N - mq|^p] \leq 2^{-p/2} m^{p/2} p^{p/2}$
.

Applying this with $m = N_T \geq \frac{n}{6d}$, the transformation from Bernoulli to $\{-1, +1\}$, and the scaling by $2\alpha - 1$, yields for $j \in S_T$, and using Eq. (22)

$$\mathbb{E}[|\mu_j - \widehat{\mu}_j|^p] \leq \left(\frac{p}{(n/6d)(2\alpha - 1)^2} \right)^{p/2}.$$

Upon summing over $j \in S_T$, we obtain

$$\mathbb{E}\left[\sum_{j \in S_T} |\mu_j - \widehat{\mu}_j|^p \right] \leq 3s \cdot \left(\frac{6(e+1)^2 d}{(e-1)^2 n\varepsilon^2} \right)^{p/2} \leq 3 \cdot 6^p \cdot s \left(\frac{pd}{n\varepsilon^2} \right)^{p/2}. \qquad \square$$

C.1.2 Communication constraints

In Algorithm 2 we propose a protocol to estimate the mean of product Bernoulli distributions under ℓ-bit communication constraints. As mentioned in the previous subsection, the ε-LDP algorithm with $\varepsilon > 1$ will follow from a simple modification of the communication-constrained one; we discuss how to privatize the latter to obtain the former at the end of the section. As in the LDP case when $\varepsilon \in (0, 1]$, in 2–10 the algorithm iteratively prunes an initial set $S_0 = [d]$ to obtain a set S_T of size $\max\{3s, \ell\}$, which denotes the set of potential non-zero coordinates. We then estimate the mean of coordinates in S_T. If $\ell > 3s$, then we can directly send the values of all coordinates in S_T and use it for estimation; otherwise, when $3s > \ell$, we again partition S_T into sets of size ℓ and each player sends the bits of its sample in this set. This is done in Lines 11–18. We state the performance of Algorithm 2 below.

Proposition 2. *Fix $p \in [1, \infty]$. For $n \geq 1$ and $\ell \leq d$, we have Algorithm 2 is an (n, γ)-estimator using $\mathcal{W}_\ell$ under ℓ_p loss for $\mathcal{B}_{d,s}$ with $\gamma = O\left(\sqrt{\frac{pds^{2/p}}{n\ell} + \frac{(p+\log(2\ell/s))s^{2/p}}{n}} \right)$ for $p \leq 2\log s$ and $\gamma = O\left(\sqrt{\frac{d\log s}{n\ell} + \frac{\log \ell}{n}} \right)$ for $p > 2\log s$.*

When $\ell \leq 3s$, the bound we get is $\gamma \lesssim \sqrt{\frac{pds^{2/p}}{n\ell}}$. The analysis is almost identical to the case under LDP constraints, since in both cases, the information we get about coordinate j are samples from a Rademacher distribution with mean $(2\alpha - 1)\mu_j$. There are only two differences. (i) $\alpha = 1$ instead of $\Theta(\varepsilon^2)$. (ii) There is a factor of ℓ more players in the corresponding groups. Combing both factors, we can obtain the desired bound by replacing ε^2 by ℓ. We omit the detailed proof in this case.

When $\ell > 3s$, after $T \asymp \log(d/\ell)$ rounds, we can find a subset S_T of size ℓ which contains most of the coordinates with large biases. The protocol then asks new players to send all coordinates within S_T using ℓ bits. In this case, it would be enough to prove Lemma 5 since for the coordinates outside S_T, we can show the error is small following exactly the same steps as the proof for bouding the first term in Lemma 4 as we explained in the case when $\ell \leq 3s$.

Lemma 5. *Let S_T be the subset obtained from the first stage of Algorithm 2, we have*

$$\mathbb{E}\left[\sum_{j \in S_T} |\mu_j - \widehat{\mu}_j|^p \right] = O\left(s \left(\frac{p + \log \frac{2\ell}{s}}{n} \right)^{p/2} \right).$$

Proof. Similar to Lemma 4, we will prove that the statement is true for any realization of S_T, which is a stronger statement than the claim.

$$\begin{aligned}
\mathbb{E}\left[\sum_{j \in S_T} |\mu_j - \widehat{\mu}_j|^p \right] &= \mathbb{E}\left[\sum_{j \in S_T} |\mu_j - \widehat{\mu}_j|^p \mathbb{1}\{j \in S_{T+1}\} \right] + \mathbb{E}\left[\sum_{j \in S_T} |\mu_j|^p \mathbb{1}\{j \notin S_{T+1}\} \right] \\
&\leq \mathbb{E}\left[\sum_{j \in S_{T+1}} |\mu_j - \widehat{\mu}_j|^p \right] + \sum_{j \in S_T} |\mu_j|^p \Pr[j \notin S_{T+1}].
\end{aligned}$$

Algorithm 2 ℓ-bit protocol for estimating product of Bernoulli family

Require: n players, dimension d, sparsity parameter s, communication bound ℓ.

Set $T := \log_3(d/\max\{3s, \ell\})$, $S_0 := [d]$, $N_0 := \frac{n\ell}{18d}$.
for $t = 1, 2, \ldots, T$ **do**
  Set $P := \frac{d}{3^{t-1}\ell}$, and partition S_{t-1} into P subsets $S_{t-1,1}, \ldots, S_{t-1,P}$, each of size ℓ.
  for $j = 1, 2, \ldots, P$ **do**
    Get a group of new players $G_{t,j}$ of size $N_t = N_0 \cdot 2^t$.
    Player $i \in G_{t,j}$, upon observing $X_i \in \{-1,+1\}^d$ sends the message $Y_i = \{(X_i)_x\}_{x \in S_{t-1,j}}$.
    For $x \in S_{t-1,j}$, let $M_{t,x} := \sum_{i \in G_{t,j}} (X_i)_x$.
  Set $S_t \subseteq S_{t-1}$ to be the set of indices with the largest $|M_{t,x}|$ and $|S_t| = |S_{t-1}|/3$.
if $\ell \leq 3s$ **then**
  Partition S_T into $3s/\ell$ subsets of size ℓ each, $S_{T,j}$, $j \in [3s/\ell]$.
  for $j = 1, \ldots, 3s/\ell$ **do**
    Get a new group $G_{T+1,j}$ of players of size $n\ell/(6s)$.
    Player $i \in G_{T+1,j}$, sends the message $Y_i = \{(X_i)_x\}_{x \in S_{T,j}}$.
    For $x \in S_{T,j}$, let $M_{T+1,x} = \sum_{i \in G_{T+1,j}} (X_i)_x$. Set

$$\widehat{\mu}_x := \frac{6s}{n\ell} M_{T+1,x},$$

    For $x \notin S_T$, set $\widehat{\mu}_x = 0$.
if $\ell > 3s$ **then**,
  Get $n/2$ new players G_{T+1} and for $i \in G_{T+1}$, player i sends $Y_i = \{(X_i)_x\}_{x \in S_T}$ ▷ This can be done since $|S_T| = \ell$ if $\ell > 3s$.
  For $x \in S_T$, let $M_{T+1,x} = \sum_{i \in G_{T+1,j}} (X_i)_x$. Set $S_{T+1} \subseteq S_T$ to be the set of indices with the largest $|M_{T+1,x}|$ and $|S_{T+1}| = 3s$. For all $x \in S_{T+1}$, set

$$\widehat{\mu}_x := \frac{2}{n} M_{T+1,x},$$

  and for all $x \notin S_{T+1}, \widehat{\mu}_x = 0$.
return $\widehat{\mu}$.

Fix S_{T+1}. For each $j \in S_{T+1}$, $M_{T+1,j}$ is binomially distributed with mean μ_j and $n/2$ trials. By similar computations as Lemma 4, we have

$$\mathbb{E}\left[\sum_{j \in S_{T+1}} |\mu_j - \widehat{\mu}_j|^p\right] = O\left(s\left(\frac{p}{n}\right)^{p/2}\right). \tag{23}$$

Next we show for all $j \in S_T$ such that $\mu_j \neq 0$,

$$|\mu_j|^p \Pr[\, j \notin S_{T+1} \,] \leq 2\left(\frac{p \vee 64 \ln \frac{2\ell}{s}}{n}\right)^{p/2}. \tag{24}$$

If $|\mu_j| \leq H' := 8\sqrt{\frac{\ln \frac{2\ell}{s}}{n}}$, Eq. (24) always holds since $\Pr[\, j \notin S \,] \leq 1$. Hence we hereafter assume that $|\mu_j| > H'$, and write $\mu_j = \beta_j H'$ for some $\beta_j > 1$.

Let $a_{T+1,j}$ be the number of coordinates j' with $\mu_{j'} = 0$ and $|M_{T+1,j'}| \geq \frac{n}{2} \cdot \frac{\beta_j H'}{2}$. Then since S_{T+1} contains the top $3s$ coordinates with the largest magnitude of the sum, we have $j \notin S_{T+1}$ happens only if at least one of the following occurs (i) $a_{T+1,j} > 2s$, or (ii) $M_{T+1,j} < \frac{n}{2} \cdot \frac{\beta_j H'}{2}$.

By Hoeffding's inequality, we have

$$\Pr\left[\, M_{T+1,j} < \frac{n}{2} \cdot \frac{\beta_j H'}{2} \,\right] \leq \exp\left(-\frac{1}{2} \cdot \frac{n}{2} \cdot \left(\frac{\beta_j H'}{2}\right)^2\right) = \left(\frac{2\ell}{s}\right)^{-4\beta_j^2} := p_{T+1,j}.$$

Similarly, for any j' such that $\mu_{j'} = 0$,

$$\Pr\left[|M_{T+1,j'}| \geq \frac{n}{2} \cdot \frac{\beta_j H'}{2} \right] \leq 2p_{T+1,j}.$$

Since all coordinates are independent, $a_{T+1,j}$ is binomially distributed with mean at most $2p_{T+1,j}\ell$, and therefore, by Markov's inequality,

$$\Pr[\, a_{T+1,j} > 2s \,] \leq \frac{2p_{T+1,j}\ell}{2s} \leq \left(\frac{2\ell}{s}\right)^{1-4\beta_j^2} \leq \left(\frac{2\ell}{s}\right)^{-3\beta_j^2}$$

the last step since $\beta_j > 1$. By a union bound, we have

$$\Pr[\, j \notin S_T \,] \leq \Pr[\, a_{T+1,j} > 2s \,] + \Pr\left[M_{T+1,j} < \frac{1}{4}\frac{n}{2} \cdot \frac{\beta_j H'}{2} \right] \leq 2\left(\frac{2\ell}{s}\right)^{-3\beta_j^2}.$$

Using the inequality $x^p a^{-x^2} \leq \left(\frac{p}{2e \ln a}\right)^{p/2}$ which holds for all $x > 0$, we get overall

$$|\mu_j|^p \cdot \Pr[\, j \notin S_T \,] \leq 2H'^p \beta_j^p \left(\frac{2\ell}{s}\right)^{-4\beta_j^2} \leq 2\left(\frac{p}{en}\right)^{p/2},$$

establishing Eq. (24). Combining Eq. (23) and Eq. (24) concludes the proof Lemma 5 since there are at most s unbiased coordinates. □

Algorithm under LDP with $\varepsilon > 1$ To get a ε-LDP algorithm in the regime $\varepsilon > 1$ (low-privacy regime), we perform the following changes to obtain a private algorithm from Algorithm 2:

- Each user independently flips each coordinate of their local sample to get Z_i where, for all $x \in [d]$, $(Z_i)_x = (X_i)_x$ with probability $\frac{e}{e+1}$ and $(Z_i)_x = 1 - (X_i)_x$ with probability $\frac{1}{e+1}$ (note that this corresponds to applying Randomized Response independently to each bit with privacy parameter 1).
- Users then follow Algorithm 2 with the setting $\ell = \lfloor \varepsilon \rfloor$ and local data $\{Z_i\}_{i\in[n]}$, and obtain estimate $\widehat{\mu}$.
- The final estimate is then $\frac{e+1}{e-1}\widehat{\mu}$.

The privacy guarantee of the algorithm comes from the fact that Algorithm 2 sends at most $\ell = \lfloor \varepsilon \rfloor$ coordinates of each Z_i, and for any S with $|S| \leq \lfloor \varepsilon \rfloor$

$$\frac{\Pr[\, \{(Z_i)_x\}_{x\in S} \mid X_i \,]}{\Pr[\, \{(Z_i)_x\}_{x\in S} \mid X_i' \,]} = \prod_{x\in S} \frac{\Pr[\, (Z_i)_x \mid (X_i)_x \,]}{\Pr[\, (Z_i)_x \mid (X_i')_x \,]} \leq e^{\lfloor \varepsilon \rfloor}.$$

The utility guarantee follows from observing that $\mu_Z = \frac{e-1}{e+1}\mu$ and hence any ℓ_p error guarantee will be preserved up to a constant.

C.2 Gaussian Mean Estimation

Recall that $\mathcal{G}_{d,s}$ denotes the family of d-dimensional spherical Gaussian distributions with s-sparse mean in $[-1,1]^d$, *i.e.*,

$$\mathcal{G}_{d,s} = \left\{\, \mathcal{G}(\mu, \mathbb{I}) \,:\, \|\mu\|_\infty \leq 1, \|\mu\|_0 \leq s \,\right\}. \tag{25}$$

We will prove the following results for LDP and communication constraints, respectively.

Proposition 3. *Fix $p \in [1,\infty]$. For $n \geq 1$ and $\varepsilon \in (0,1]$, there exists an (n,γ)-estimator using $\mathcal{W}_\varepsilon$ under ℓ_p loss for $\mathcal{G}_{d,s}$ with $\gamma = O\left(\sqrt{\frac{pds^{2/p}}{n\varepsilon^2}}\right)$ for $p \leq 2\log s$ and $\gamma = O\left(\sqrt{\frac{d\log s}{n\varepsilon^2}}\right)$ for $p > 2\log s$.*

Proposition 4. *Fix $p \in [1,\infty]$. For $n \geq 1$ and $\ell \leq d$, there exists an (n,γ)-estimator using $\mathcal{W}_\ell$ under ℓ_p loss for $\mathcal{G}_{d,s}$ with $\gamma = O\left(\sqrt{\frac{pds^{2/p}}{n\ell} + \frac{(p+\log(2\ell/s))s^{2/p}}{n}}\right)$ for $p \leq 2\log s$ and $\gamma = O\left(\sqrt{\frac{d\log s}{n\ell} + \frac{\log \ell}{n}}\right)$ for $p > 2\log s$.*

We reduce the problem of Gaussian mean estimation to that of Bernoulli mean estimation and then invoke Propositions 1 and 2 from the previous section. At the heart of the reduction is a simple idea that was used in, *e.g.*, [10, 2, 11]: the sign of a Gaussian random variable already preserves sufficient information about the mean. Details follow.

Let $\mathbf{p} \in \mathcal{G}_{d,s}$ with mean $\mu(\mathbf{p}) = (\mu(\mathbf{p})_1, \dots, \mu(\mathbf{p})_d)$. For $X \sim \mathbf{p}$, let $Y = (\operatorname{sign}(X_i))_{i\in[d]} \in \{-1,+1\}^d$ be a random variable indicating the signs of the d coordinates of X. By the independence of the coordinates of X, note that Y is distributed as a product Bernoulli distribution (in $\mathcal{B}_d$) with mean vector $\nu(\mathbf{p})$ given by

$$\nu(\mathbf{p})_i = 2 \Pr_{X\sim\mathbf{p}}[\, X_i > 0\,] - 1 = \operatorname{Erf}\left(\frac{\mu(\mathbf{p})_i}{\sqrt{2}}\right), \qquad i \in [d], \tag{26}$$

and, since $|\mu(\mathbf{p})_i| \le 1$, we have $\nu(\mathbf{p}) \in [-\eta,\eta]^d$, where $\eta := \operatorname{Erf}\big(1/\sqrt{2}\big) \approx 0.623$. Moreover, it is immediate to see that each player, given a sample from $\mathbf{p}$, can convert it to a sample from the corresponding product Bernoulli distribution. We now show that a good estimate for $\nu(\mathbf{p})$ yields a good estimate for $\mu(\mathbf{p})$.

Lemma 6. *Fix any $p \in [1,\infty)$, and $\mathbf{p} \in \mathcal{G}_d$. For $\widehat{\nu} \in [-\eta,\eta]^d$, define $\widehat{\mu} \in [-1,1]^d$ by $\widehat{\mu}_i := \sqrt{2}\operatorname{Erf}^{-1}(\widehat{\nu}_i)$, for all $i \in [d]$. Then*

$$\|\mu(\mathbf{p}) - \widehat{\mu}\|_p \le \sqrt{\frac{e\pi}{2}} \cdot \|\nu(\mathbf{p}) - \widehat{\nu}\|_p \,.$$

Proof. By computing the maximum of its derivative,[7] we observe that the function Erf^{-1} is $\frac{\sqrt{e\pi}}{2}$-Lipschitz on $[-\eta,\eta]$. By the definition of $\widehat{\mu}$ and recalling Eq. (26), we then have

$$\|\mu(\mathbf{p}) - \widehat{\mu}\|_p^p = \sum_{i=1}^d |\mu(\mathbf{p})_i - \widehat{\mu}_i|^p = 2^{p/2} \cdot \sum_{i=1}^d \left|\operatorname{Erf}^{-1}(\nu_i) - \operatorname{Erf}^{-1}(\widehat{\nu}_i)\right|^p \le \left(\frac{e\pi}{2}\right)^{p/2} \cdot \sum_{i=1}^d |\nu_i - \widehat{\nu}_i|^p,$$

where we used the fact that $\nu, \widehat{\nu} \in [-\eta,\eta]^d$. □

As previously discussed, combining Lemma 6 with Propositions 1 and 2 (with $\gamma' := \sqrt{\frac{2}{e\pi}}\gamma$) immediately implies Propositions 3 and 4 for $p \in [1,\infty]$.

Remark 3. Note that for the Gaussian family, we also consider the linear measurement constraint. Under linear measurement constraints, we can use the linear measurement matrix to obtain r out of d coordinates and perform the above reduction to product of Bernoulli family. The obtained bound will be same as that under communication constraints.

D Relation to other lower bound methods

We now discuss how our techniques compare with other existing approaches for proving lower bounds under information constraints. Specifically, we clarify the relationship between our technique and the approach using strong data processing inequalities (SDPI) as well as that based on van Trees inequality (a generalization of the Cramér–Rao bound).

D.1 Strong data processing inequalities

We note first that the bound in Eq. (5) can be interpreted as a strong data processing inequality. Indeed, the average discrepancy on the left-side of inequality can be viewed as the average information Y^n reveals about each bit of Z. Here the information is measured in terms of total variation distance. The information quantity on the right-side denotes the information between the input X^n and the output Y^n of the channels. Since the Markov relation $Z^n — X^n — Y^n$ holds, the inequality is thus a strong data processing inequality with strong data processing constant roughly σ^2/k. Such

[7] Specifically, we have that $\max_{x\in[-\eta,\eta]} \operatorname{Erf}^{-1}(x) = 1/\sqrt{2}$ by definition of η and monotonicity of Erf. Recalling then that, for all $x \in [-\eta,\eta]$, $(\operatorname{Erf}^{-1})'(x) = \frac{1}{\operatorname{Erf}'(\operatorname{Erf}^{-1}(x))} = \frac{\sqrt{\pi}}{2} e^{(\operatorname{Erf}^{-1}(x))^2} \le \frac{\sqrt{\pi}}{2} e^{\frac{1}{2}}$, we get the Lipschitzness claim.

strong data processing inequalities were used to derive lower bounds for statistical estimation under communication constraints in [34, 10, 31]. We note that our approach recovers these bounds, and further applies to arbitrary constraints captured by $\mathcal{W}$.

D.2 Connection to the van Trees inequality

The average information bound in (3), in fact, allows us to recover bounds similar to the van Trees inequality-based bounds developed in [7] and [8].

For $\Theta \subset \mathbb{R}^k$ and a parametric family[8] $\mathcal{P}_\Theta = \{\mathbf{p}_\theta, \theta \in \Theta\}$, recall that the Fisher information matrix $J(\theta)$ is a $k \times k$ matrix given by, under some mild regularity conditions,

$$J(\theta)_{i,j} = -\mathbb{E}_{\mathbf{p}_\theta}\left[\frac{\partial^2 \log \mathbf{p}_\theta}{\partial\theta_i \partial\theta_j}(X)\right], \quad i, j \in [k].$$

In particular, the diagonal entries equal

$$J(\theta)_{i,i} = \mathbb{E}_{\mathbf{p}_\theta}\left[\left(\frac{1}{\mathbf{p}_\theta(X)} \cdot \frac{\partial \mathbf{p}_\theta}{\partial\theta_i}(X)\right)^2\right], \quad i \in [k].$$

For our application, given a channel $W \in \mathcal{W}$, we consider the family $\mathcal{P}_\Theta^W := \{\mathbf{p}_\theta^W, \theta \in \Theta\}$ of distributions induced on the output of the channel W when the input distributions are from $\mathcal{P}_\Theta$. We denote the Fisher information matrix for this family by $J^W(\theta)$, which we compute next under a refined version of our Assumption 1 described below.

Let θ be a point in the interior of Θ and $\mathbf{p}_\theta$ be differentiable at θ. We set $\theta_z := \theta + \frac{\gamma}{2} z$, $z \in \{-1, +1\}^k$, and make the following assumption about the structure of the parametric family of distribution: For all $z \in \{-1, +1\}^k$ and $i \in [k]$,

$$\frac{\mathrm{d}\mathbf{p}_{z^{\oplus i}}}{\mathrm{d}\mathbf{p}_z} = 1 + \gamma \xi_{z,i}^\gamma + \gamma^2 \psi_{z,i}^\gamma, \tag{27}$$

where $\mathbb{E}_{\mathbf{p}_z}\left[\xi_{z,i}^\gamma(X)^2\right]$ and $\mathbb{E}_{\mathbf{p}_z}\left[\psi_{z,i}^\gamma(X)^2\right]$ are assumed to be uniformly bounded for γ sufficiently small; for concreteness, we assume $\mathbb{E}_{\mathbf{p}_z}\left[\psi_{z,i}^\gamma(X)^2\right] \le c^2$ for a constant c, for all γ sufficiently small. Let $\xi_{z,i}(x) := \lim_{\gamma \to \mathbf{0}} \xi_{z,i}^\gamma(x)$, for all x.

In applications, we expect the dependence of $\xi_{z,i}^\gamma$ on γ to be "mild," and, in essence, the assumption above provides a linear expansion of the term $\alpha_{z,i}\phi_{z,i}$ from Assumption 1 as a function of the perturbation parameter γ. Assuming that the densities are differentiable as a function of θ, for the distribution $\mathbf{p}_\theta^W$ of the output of a channel W with input $X \sim \mathbf{p}_\theta$, we get

$$\begin{aligned}
\frac{\partial \mathbf{p}_\theta^W(y)}{\partial \theta_i} &= z_i \lim_{\gamma \to 0} \frac{\mathbf{p}_{\theta_z}^W(y) - \mathbf{p}_{\theta_{z^{\oplus i}}}^W(y)}{\gamma} \\
&= z_i \lim_{\gamma \to 0} \mathbb{E}_{\mathbf{p}_z}\left[(\xi_{z,i}^\gamma(X) + \gamma\psi_{z,i}^\gamma(X)) W(y \mid X)\right] \\
&= z_i \mathbb{E}_{\mathbf{p}_\theta}[\xi_{z,i} W(y \mid X)],
\end{aligned}$$

where we used Eq. (27), the fact that $\lim_{\gamma \to 0} \theta_z = \theta$, the fact that $\mathbb{E}_{\mathbf{p}_z}\left[\psi_{z,i}^\gamma(X) W(y \mid X)\right] \le c\sqrt{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)^2]} \le c$, and the dominated convergence theorem. Thus, we get

$$\mathrm{Tr}\big(J^W(\theta)\big) = \sum_{i=1}^k \int_{\mathcal{Y}} \frac{\mathbb{E}_{\mathbf{p}_\theta}[\xi_{z,i}(X) W(y \mid X)]^2}{\mathbb{E}_{\mathbf{p}_\theta}[W(y \mid X)]} \,\mathrm{d}\mu\,. \tag{28}$$

Our information contraction bound will be seen later (Section 5) to yield lower bounds for expected estimation error. For concreteness, we give a preview of a version here. We assume for simplicity that $\mathcal{W}_t = \mathcal{W}$ for all t and consider the ℓ_2 loss function for the dense ($\tau = 1/2$) case. By following

[8] We assume that each distribution $\mathbf{p}_\theta$ has a density with respect to a common measure ν, and, with a slight abuse of notation, denote the density of $\mathbf{p}_\theta$ also by $\mathbf{p}_\theta(X)$.

the proof of Lemma 1 below, given an (n,γ)-estimator $\hat{\theta} = \hat{\theta}(Y^n, U)$ of $\mathcal{P}_\Theta$ using $\mathcal{W}^n$ under ℓ_2 loss, we can find an estimator $\hat{Z} = \hat{Z}(Y^n, U)$ such that

$$\gamma^2 \sum_{i=1}^{k} \Pr\Big[\, \hat{Z}_i \neq Z_i \,\Big] = \mathbb{E}\Big[\big\|\theta_Z - \theta_{\hat{Z}}\big\|_2^2\Big] \leq 4\gamma^2,$$

whereby

$$\frac{1}{k}\sum_{i=1}^{k} \mathrm{d}_{\mathrm{TV}}\Big(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n}\Big) \geq 1 - \frac{2}{k}\sum_{i=1}^{k}\Pr\Big[\, \hat{Z}_i \neq Z_i \,\Big] \geq 1 - \frac{8\gamma^2}{k\gamma^2}.$$

Upon setting $\gamma := 4\gamma/\sqrt{k}$, we get that the left-side of Eq. (3) is bounded below by $1/4$. For the same γ and under Eq. (27), the right-side evaluates to

$$\begin{aligned}
&\frac{4\gamma^2 n}{k} \max_{z\in\mathcal{Z}} \max_{W\in\mathcal{W}} \sum_{i=1}^{k} \int_{\mathcal{Y}} \frac{\mathbb{E}_{\mathbf{p}_z}\big[(\xi_{z,i}^{\gamma}(X) + \gamma\psi_{z,i}^{\gamma}(X))W(y \mid X)\big]^2}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]}\, \mathrm{d}\mu \\
&\leq \frac{8\gamma^2 n}{k} \max_{z\in\mathcal{Z}} \max_{W\in\mathcal{W}} \sum_{i=1}^{k} \int_{\mathcal{Y}} \frac{\mathbb{E}_{\mathbf{p}_z}\big[\xi_{z,i}^{\gamma}(X)W(y \mid X)\big]^2 + \gamma^2\mathbb{E}_{\mathbf{p}_z}\big[\psi_{z,i}^{\gamma}(X)W(y \mid X)\big]^2}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]}\, \mathrm{d}\mu \\
&\leq \frac{128\gamma^2 n}{k^2}\left(\max_{z\in\mathcal{Z}} \max_{W\in\mathcal{W}} \sum_{i=1}^{k} \int_{\mathcal{Y}} \frac{\mathbb{E}_{\mathbf{p}_z}\big[\xi_{z,i}^{\gamma}(X)W(y \mid X)\big]^2}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]}\, \mathrm{d}\mu + c^2\gamma^2\right),
\end{aligned}$$

where we used $(a+b)^2 \leq 2(a^2+b^2)$ and

$$\int_{\mathcal{Y}} \frac{\mathbb{E}_{\mathbf{p}_z}\big[\psi_{z,i}^{\gamma}(X)W(y \mid X)\big]^2}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]}\, \mathrm{d}\mu \leq \int_{\mathcal{Y}} \mathbb{E}_{\mathbf{p}_z}\big[\psi_{z,i}^{\gamma}(X)^2 W(y \mid X)\big]\, \mathrm{d}\mu = \mathbb{E}_{\mathbf{p}_z}\big[\psi_{z,i}^{\gamma}(X)^2\big] \leq c^2.$$

Therefore, Eq. (3) yields

$$\gamma^2 \geq \frac{k^2}{256 \cdot n\left(\max_{z\in\mathcal{Z}}\max_{W\in\mathcal{W}}\sum_{i=1}^{k}\int_{\mathcal{Y}} \frac{\mathbb{E}_{\mathbf{p}_z}\left[\xi_{z,i}^{\gamma}(X)W(y|X)\right]^2}{\mathbb{E}_{\mathbf{p}_z}[W(y|X)]}\, \mathrm{d}\mu + c^2\right)}.$$

This bound is, in effect, the same as the van Trees inequality with $\mathrm{Tr}\big(J^W(\theta)\big)$ replaced by

$$g(\gamma) := \sum_{i=1}^{k} \int_{\mathcal{Y}} \frac{\mathbb{E}_{\mathbf{p}_z}[\phi_{z,i}(X)W(y \mid X)]^2}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]}\, \mathrm{d}\mu\,.$$

In fact, in view of Eq. (28), $\mathrm{Tr}\big(J^W(\theta)\big) = \lim_{\gamma\to 0} g(\gamma) =: g(0)$. Thus, our general lower bound will recover van Trees inequality-based bounds when Eq. (27) holds and $g(\gamma) \approx g(0)$. We note that Eq. (27) holds for all the families considered in this paper (see Eq. (37) for product Bernoulli, Eq. (42) for Gaussian, and Eq. (50) for discrete distributions). We close this discussion by noting that results in Section 3 are obtained by deriving bounds for $g(\gamma)$ which apply for all γ and, therefore, also for $g(0) = \mathrm{Tr}\big(J^W(\theta)\big)$.

E Missing proofs in Section 3

E.1 Proof of Theorem 1

Consider $Z = (Z_1, \ldots, Z_k) \in \{-1, 1\}^k$ where $Z_1, \ldots, Z_k$ are i.i.d. with $\Pr[\, Z_i = 1\,] = \tau$. For a fixed $i \in [k]$, let

$$\begin{aligned}
\mathbf{p}_{+i}^{Y^n} &:= \mathbb{E}_Z\Big[\mathbf{p}_Z^{Y^n} \mid Z_i = +1\Big] = \sum_{z:z_i=+1}\Big(\prod_{j\neq i}\tau^{\frac{1+z_j}{2}}(1-\tau)^{\frac{1-z_j}{2}}\Big)\mathbf{p}_z^{Y^n} \\
\mathbf{p}_{-i}^{Y^n} &:= \mathbb{E}_Z\Big[\mathbf{p}_Z^{Y^n} \mid Z_i = -1\Big] = \sum_{z:z_i=-1}\Big(\prod_{j\neq i}\tau^{\frac{1+z_j}{2}}(1-\tau)^{\frac{1-z_j}{2}}\Big)\mathbf{p}_z^{Y^n},
\end{aligned}$$

the partial mixtures of message distributions conditioned on Z_i. We will rely on the following lemma, which relates the desired average discrepancy between the $\mathbf{p}_{+i}^{Y^n}$ and $\mathbf{p}_{-i}^{Y^n}$'s to the sum of n "local" discrepancy measures (in the form of Hellinger distances between local messages). Each local measure can then be easily bounded in terms of the density $\mathbf{p}_z$ and the channel W to get the desired bound.

Lemma 7. *With the notation of Theorem 1, we have*

$$\left(\frac{1}{k}\sum_{i=1}^{k} \mathrm{d_{TV}}\left(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n}\right)\right)^2 \leq \frac{14}{k}\sum_{t=1}^{n} \max_{z\in\mathcal{Z}} \max_{W\in\mathcal{W}_t} \sum_{i=1}^{k} \mathrm{d_H}\left(\mathbf{p}_z^W, \mathbf{p}_{z\oplus i}^W\right)^2, \tag{29}$$

where $\mathbf{p}_z^W$ denotes the distribution of $Y \sim W(\cdot \mid X)$ when $X \sim \mathbf{p}_z$.

The proof of the lemma is rather involved and constitutes the core of the argument. We defer it to the end of the section and show first how it implies Theorem 1. For all z and W, we have

$$\begin{aligned}
\mathrm{d_H}\left(\mathbf{p}_z^W, \mathbf{p}_{z\oplus i}^W\right)^2 &= \frac{1}{2}\int_{y\in\mathcal{Y}} \left(\sqrt{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]} - \sqrt{\mathbb{E}_{\mathbf{p}_{z\oplus i}}[W(y \mid X)]}\right)^2 \mathrm{d}\mu \\
&= \frac{1}{2}\int_{\mathcal{Y}} \left(\frac{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)] - \mathbb{E}_{\mathbf{p}_{z\oplus i}}[W(y \mid X)]}{\sqrt{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]} + \sqrt{\mathbb{E}_{\mathbf{p}_{z\oplus i}}[W(y \mid X)]}}\right)^2 \mathrm{d}\mu \\
&\leq \frac{1}{2}\int_{\mathcal{Y}} \frac{(\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)] - \mathbb{E}_{\mathbf{p}_{z\oplus i}}[W(y \mid X)])^2}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]} \mathrm{d}\mu\,.
\end{aligned} \tag{30}$$

Moreover, under Assumption 1; for any $W \in \mathcal{W}_t$ and $y \in \mathcal{Y}$,

$$\mathbb{E}_{\mathbf{p}_{z\oplus i}}[W(y \mid X)] = \mathbb{E}_{\mathbf{p}_z}\left[\frac{\mathrm{d}\mathbf{p}_{z\oplus i}}{\mathrm{d}\mathbf{p}_z} \cdot W(y \mid X)\right] = \mathbb{E}_{\mathbf{p}_z}[(1+\phi_{z,i}(X)) \cdot W(y \mid X)]\,.$$

Plugging this back into (30), we get

$$\mathrm{d_H}\left(\mathbf{p}_z^W, \mathbf{p}_{z\oplus i}^W\right)^2 \leq \frac{1}{2}\int_{\mathcal{Y}} \frac{\mathbb{E}_{\mathbf{p}_z}[\phi_{z,i}(X)W(y \mid X)]^2}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]} \mathrm{d}\mu\,.$$

Combining this with Lemma 7 concludes the proof of Theorem 1.

Proof of Lemma 7. Our first step is to use the Cauchy–Schwarz inequality, followed by an inequality relating total variation and Hellinger distances:

$$\begin{aligned}
\frac{1}{k}\left(\sum_{i=1}^{k} \mathrm{d_{TV}}\left(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n}\right)\right)^2 &\leq \sum_{i=1}^{k} \mathrm{d_{TV}}\left(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n}\right)^2 \\
&\leq 2\sum_{i=1}^{k} \mathrm{d_H}\left(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n}\right)^2 \\
&\leq 2\sum_{i=1}^{k} \mathbb{E}_Z\left[\mathrm{d_H}\left(\mathbf{p}_{Z}^{Y^n}, \mathbf{p}_{Z\oplus i}^{Y^n}\right)^2 \mid Z_i = +1\right] \\
&= 2\sum_{i=1}^{k} \mathbb{E}_Z\left[\mathrm{d_H}\left(\mathbf{p}_{Z}^{Y^n}, \mathbf{p}_{Z\oplus i}^{Y^n}\right)^2\right],
\end{aligned} \tag{31}$$

where the last inequality uses joint convexity of squared Hellinger distance, and the final identity is due to independence of each coordinate of Z and symmetry of Hellinger whereby $\mathbb{E}_Z\left[\mathrm{d_H}\left(\mathbf{p}_{Z}^{Y^n}, \mathbf{p}_{Z\oplus i}^{Y^n}\right)^2 \mid Z_i = +1\right] = \mathbb{E}_Z\left[\mathrm{d_H}\left(\mathbf{p}_{Z}^{Y^n}, \mathbf{p}_{Z\oplus i}^{Y^n}\right)^2 \mid Z_i = -1\right]$.

In order to bound the resulting terms of the sum, we will rely on the so-called *cut-paste* property of Hellinger distance [6]. Before doing so, we will require an additional piece of notation: for fixed $z \in \mathcal{Z}$, $i \in [k]$, $t \in [n]$, let $\mathbf{p}_{t\leftarrow z\oplus i}^{Y^n}$ denote the message distribution where player t gets a sample from

$\mathbf{p}_{z\oplus i}$ and all other players get samples from $\mathbf{p}_z$. That is, for all $y^n \in \mathcal{Y}^n$, the density of $\mathbf{p}^{Y^n}_{t\leftarrow z\oplus i}$ with respect to the underlying product measure $\mu^{\otimes n}$ is given by

$$\frac{\mathrm{d}\mathbf{p}^{Y^n}_{t\leftarrow z\oplus i}}{\mathrm{d}\mu^{\otimes n}}(y^n) = \mathbb{E}_{X_t\sim\mathbf{p}_{z\oplus i}}\left[W^{y^{t-1}}(y_t \mid X_t)\right] \cdot \prod_{j\neq t} \mathbb{E}_{X_j\sim\mathbf{p}_z}\left[W^{y^{j-1}}(y_j \mid X_j)\right]. \tag{32}$$

The following lemma, due to [22], allows us to relate $\mathrm{d}_{\mathrm{H}}\big(\mathbf{p}^{Y^n}_z, \mathbf{p}^{Y^n}_{z\oplus i}\big)$, the distance between message distributions when all players get observations from $\mathbf{p}_z$, or all from $\mathbf{p}_{z\oplus i}$, to the distances $\mathrm{d}_{\mathrm{H}}\big(\mathbf{p}^{Y^n}_z, \mathbf{p}^{Y^n}_{t\leftarrow z\oplus i}\big)$ where only *one* of the n players gets a sample from $\mathbf{p}_{z\oplus i}$.

Lemma 8 ([22, Theorem 7])**.** *There exists $c_{\mathrm{H}} > 0$ such that for all $z \in \mathcal{Z}$ and $i \in [k]$,*

$$\mathrm{d}_{\mathrm{H}}\left(\mathbf{p}^{Y^n}_z, \mathbf{p}^{Y^n}_{z\oplus i}\right)^2 \leq c_{\mathrm{H}} \sum_{t=1}^{n} \mathrm{d}_{\mathrm{H}}\left(\mathbf{p}^{Y^n}_z, \mathbf{p}^{Y^n}_{t\leftarrow z\oplus i}\right)^2.$$

Moreover, one can take $c_{\mathrm{H}} = 2\prod_{t=1}^{\infty} \frac{1}{1-2^{-t}} < 7$.

Combining Eq. (31) and Lemma 8, we get

$$\begin{aligned}\frac{1}{k}\left(\sum_{i=1}^{k} \mathrm{d}_{\mathrm{TV}}\left(\mathbf{p}^{Y^n}_{+i}, \mathbf{p}^{Y^n}_{-i}\right)\right)^2 &\leq 14 \sum_{i=1}^{k}\sum_{t=1}^{n} \mathbb{E}_Z\left[\mathrm{d}_{\mathrm{H}}\left(\mathbf{p}^{Y^n}_Z, \mathbf{p}^{Y^n}_{t\leftarrow Z\oplus i}\right)^2\right] \\ &= 14 \sum_{t=1}^{n} \mathbb{E}_Z\left[\sum_{i=1}^{k} \mathrm{d}_{\mathrm{H}}\left(\mathbf{p}^{Y^n}_Z, \mathbf{p}^{Y^n}_{t\leftarrow Z\oplus i}\right)^2\right].\end{aligned} \tag{33}$$

In view of bounding the RHS of (33) term by term, fix $j \in [n]$ and $z \in \mathcal{Z}$. Recalling the expression of $\mathbf{p}^{Y^n}_{t\leftarrow z\oplus i}$ from (32), unrolling the definition of Hellinger distance, and recalling (32), we have

$$\begin{aligned}
&2\sum_{i=1}^{k} \mathrm{d}_{\mathrm{H}}\left(\mathbf{p}^{Y^n}_z, \mathbf{p}^{Y^n}_{t\leftarrow z\oplus i}\right)^2 \\
&= \sum_{i=1}^{k}\int_{\mathcal{Y}^n}\left(\sqrt{\frac{\mathrm{d}\mathbf{p}^{Y^n}_z}{\mathrm{d}\mu^{\otimes n}}} - \sqrt{\frac{\mathrm{d}\mathbf{p}^{Y^n}_{t\leftarrow z\oplus i}}{\mathrm{d}\mu^{\otimes n}}}\right)^2 \mathrm{d}\mu^{\otimes n} \\
&= \sum_{i=1}^{k}\int_{\mathcal{Y}^n}\prod_{j\neq t}\mathbb{E}_{\mathbf{p}_z}\left[W^{y^{j-1}}(y_j \mid X)\right]\underbrace{\left(\sqrt{\mathbb{E}_{\mathbf{p}_z}\left[W^{y^{t-1}}(y_t \mid X)\right]} - \sqrt{\mathbb{E}_{\mathbf{p}_{z\oplus i}}\left[W^{y^{t-1}}(y_t \mid X)\right]}\right)^2}_{:=f_{i,t}(y^{t-1},y_t)}\mathrm{d}\mu^{\otimes n} \\
&= \sum_{i=1}^{k}\int_{\mathcal{Y}^{t-1}}\prod_{j<t}\mathbb{E}_{\mathbf{p}_z}\left[W^{y^{j-1}}(y_j \mid X)\right]\int_{\mathcal{Y}} f_{i,t}(y^{t-1},y_t)\int_{\mathcal{Y}^{n-t}}\prod_{j>t}\mathbb{E}_{\mathbf{p}_z}\left[W^{y^{j-1}}(y_j \mid X)\right]\mathrm{d}\mu^{\otimes(t-1)}\,\mathrm{d}\mu\,\mathrm{d}\mu^{\otimes(n-t)} \\
&= \sum_{i=1}^{k}\int_{\mathcal{Y}^{t-1}}\prod_{j<t}\mathbb{E}_{\mathbf{p}_z}\left[W^{y^{j-1}}(y_j \mid X)\right]\int_{\mathcal{Y}} f_{i,t}(y^{t-1},y_t)\Big(\int_{\mathcal{Y}^{n-t}}\prod_{j>t}\mathbb{E}_{\mathbf{p}_z}\left[W^{y^{j-1}}(y_j \mid X)\right]\mathrm{d}\mu^{\otimes(n-t)}\Big)\,\mathrm{d}\mu^{\otimes(t-1)}\,\mathrm{d}\mu \\
&= \sum_{i=1}^{k}\int_{\mathcal{Y}^{t-1}}\prod_{j<t}\mathbb{E}_{\mathbf{p}_z}\left[W^{y^{j-1}}(y_j \mid X)\right]\int_{\mathcal{Y}} f_{i,t}(y^{t-1},y_t)\,\mathrm{d}\mu\,\mathrm{d}\mu^{\otimes(t-1)} \\
&= \int_{\mathcal{Y}^{t-1}}\prod_{j<t}\mathbb{E}_{\mathbf{p}_z}\left[W^{y^{j-1}}(y_j \mid X)\right]\sum_{i=1}^{k}\int_{\mathcal{Y}} f_{i,t}(y^{t-1},y_t)\,\mathrm{d}\mu\,\mathrm{d}\mu^{\otimes(t-1)}\ ,
\end{aligned}$$

where the second-to-last identity uses the observation that, for any fixed $y^t \in \mathcal{Y}^t$,

$$\int_{\mathcal{Y}^{n-t}}\prod_{j>t}\mathbb{E}_{\mathbf{p}_z}\left[W^{y^{j-1}}(y_j \mid X)\right]\mathrm{d}\mu^{\otimes(n-t)} = 1,$$

which in turn follows upon taking marginal integrals for each coordinate. We then get from the pointwise inequality $\sum_{i=1}^{k} \int_{\mathcal{Y}^{t-1}} f_{i,t}(y^{t-1}, y_t)\,\mathrm{d}\mu \le \sup_{y' \in \mathcal{Y}^{t-1}} \sum_{i=1}^{k} \int_{\mathcal{Y}} f_{i,t}(y', y_t)\,\mathrm{d}\mu$ that

$$
\begin{aligned}
2\sum_{i=1}^{k} \mathrm{d}_{\mathrm{H}}\left(\mathbf{p}_z^{Y^n}, \mathbf{p}_{t\leftarrow z\oplus i}^{Y^n}\right)^2 &\le \int_{\mathcal{Y}^{t-1}} \prod_{j<t} \mathbb{E}_{\mathbf{p}_z}\left[W^{y^{j-1}}(y_j \mid X)\right] \sup_{y' \in \mathcal{Y}^{t-1}} \sum_{i=1}^{k} \left(\int_{\mathcal{Y}} f_{i,t}(y', y_t)\,\mathrm{d}\mu \right) \mathrm{d}\mu^{\otimes(t-1)} \\
&= \left(\sup_{y' \in \mathcal{Y}^{t-1}} \sum_{i=1}^{k} \int_{\mathcal{Y}} f_{i,t}(y', y_t)\,\mathrm{d}\mu \right) \int_{\mathcal{Y}^{t-1}} \prod_{j<t} \mathbb{E}_{\mathbf{p}_z}\left[W^{y^{j-1}}(y_j \mid X)\right] \mathrm{d}\mu^{\otimes(t-1)} \\
&= \sup_{y' \in \mathcal{Y}^{t-1}} \sum_{i=1}^{k} \int_{\mathcal{Y}} \left(\sqrt{\mathbb{E}_{\mathbf{p}_z}[W^{y'}(y \mid X)]} - \sqrt{\mathbb{E}_{\mathbf{p}_{z\oplus i}}[W^{y'}(y \mid X)]} \right)^2 \mathrm{d}\mu \\
&\le \sup_{W \in \mathcal{W}_t} \sum_{i=1}^{k} \int_{\mathcal{Y}} \left(\sqrt{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]} - \sqrt{\mathbb{E}_{\mathbf{p}_{z\oplus i}}[W(y \mid X)]} \right)^2 \mathrm{d}\mu \\
&= 2 \cdot \sup_{W \in \mathcal{W}_t} \sum_{i=1}^{k} \mathrm{d}_{\mathrm{H}}\left(\mathbf{p}_z^W, \mathbf{p}_{z\oplus i}^W\right)^2. \qquad (34)
\end{aligned}
$$

the second identity follows upon taking marginal integrals, and by replacing $f_{i,t}$ by its definition; and the second inequality using that $\left\{ W^{y'} : y' \in \mathcal{Y}^{t-1} \right\} \subseteq \mathcal{W}_t$, so that we are taking a supremum over a larger set.

Plugging this back into (33) and upper bounding the inner expectation by a maximum concludes the proof of the lemma. □

E.2 Proof of Theorem 2

Our starting point is Eq. (3) which holds under Assumption 1. We will bound the right-hand-side of Eq. (3) under assumptions of orthogonality and subgaussianity to prove the two bounds in Theorem 2.

First, under orthogonality (Assumption 2), we apply Bessel's inequality to Eq. (3). For a fixed $z \in \mathcal{Z}$, write $\psi_{z,i} = \frac{\phi_{z,i}}{\sqrt{\mathbb{E}_{\mathbf{p}_z}[\phi_{z,i}^2]}}$, and complete $(1, \psi_{z,1}, \dots, \psi_{z,k})$ to get an orthonormal basis $\mathcal{B}$ for $L^2(\mathcal{X}, \mathbf{p}_z)$. Fix any $W \in \mathcal{W}$ and $y \in \mathcal{Y}$, and, for brevity, define $a \colon \mathcal{X} \to \mathbb{R}$ as $a(x) = W(y \mid x)$. Then, we have

$$
\begin{aligned}
\sum_{i=1}^{k} \mathbb{E}[\phi_{z,i}(X)a(X)]^2 \le \alpha^2 \sum_{i=1}^{k} \mathbb{E}[\psi_{z,i}(X)a(X)]^2 = \alpha^2 \sum_{i=1}^{k} \langle a, \psi_{z,i} \rangle^2 = \alpha^2 \sum_{i=1}^{k} \langle a - \mathbb{E}[a], \psi_{z,i} \rangle^2 \\
\le \alpha^2 \sum_{\psi \in \mathcal{B}} \langle a - \mathbb{E}[a], \psi \rangle^2 = \alpha^2\, \mathrm{Var}[a(X)],
\end{aligned}
$$

where for the second identity we used the assumption that $\langle \mathbb{E}[a], \psi_{z,i} \rangle = 0$ for all $i \in [k]$ (since 1 and $\psi_{z,i}$ are orthogonal). This establishes Eq. (4).

Turning to Eq. (5), suppose that Assumption 3 holds. Fix $z \in \mathcal{Z}$, and consider any $W \in \mathcal{W}$ and $y \in \mathcal{Y}$. Upon applying Lemma 4 of the Supplement (See Supplement (Appendix B) for the precise statement and proof) to the σ^2-subgaussian random vector $\phi_z(X)$ and with $a(x)$ set to $W(y \mid x) \in [0, 1]$, we get that

$$
\begin{aligned}
\sum_{i=1}^{k} \mathbb{E}_{\mathbf{p}_z}[\phi_{z,i}(X)W(y \mid X)]^2 &= \|\mathbb{E}_{\mathbf{p}_z}[\phi_z(X)W(y \mid X)]\|_2^2 \\
&\le 2\sigma^2 \mathbb{E}_{\mathbf{p}_z}[W(y \mid X)] \cdot \mathbb{E}_{\mathbf{p}_z}\left[W(y \mid X) \log \frac{W(y \mid X)}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]} \right]
\end{aligned}
$$

Integrating over $y \in \mathcal{Y}$, this gives

$$\int_{\mathcal{Y}} \frac{\sum_{i=1}^{k} \mathbb{E}_{\mathbf{p}_z}[\phi_{z,i}(X)W(y \mid X)]^2}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]} \,\mathrm{d}\mu \le 2\sigma^2 \cdot \int_{\mathcal{Y}} \mathbb{E}_{\mathbf{p}_z}\left[W(y \mid X) \log \frac{W(y \mid X)}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]}\right] \mathrm{d}\mu$$
$$= 2\sigma^2 I(\mathbf{p}_z; W),$$

which yields the claimed bound.

E.3 Proof of Corollary 1

For any $W \in \mathcal{W}^{\text{priv},\varepsilon}$, the ε-LDP condition from Eq. (2) can be seen to imply that, for every $y \in \mathcal{Y}$,

$$W(y \mid x_1) - W(y \mid x_2) \le (e^\varepsilon - 1)W(y \mid x_3), \qquad \forall x_1, x_2, x_3 \in \mathcal{X}\,.$$

By taking expectation over x_3 then again either over x_1 or x_2 (all distributed according to $\mathbf{p}_z$), this yields

$$|W(y \mid x) - \mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]| \le (e^\varepsilon - 1)\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)], \qquad \forall x \in \mathcal{X}\,.$$

Squaring and taking the expectation on both sides, we obtain

$$\operatorname{Var}_{\mathbf{p}_z}[W(y \mid X)] \le (e^\varepsilon - 1)^2\, \mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]^2.$$

Dividing by $\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]$, summing over $y \in \mathcal{Y}$, and using $\int_{\mathcal{Y}} \mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]\,\mathrm{d}\mu = 1$ gives

$$\int_{\mathcal{Y}} \frac{\operatorname{Var}_{\mathbf{p}_z}[W(y \mid X)]}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]}\,\mathrm{d}\mu \le (e^\varepsilon - 1)^2 \int_{\mathcal{Y}} \mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]\,\mathrm{d}\mu = (e^\varepsilon - 1)^2,$$

thus establishing (6). For the bound of e^ε, observe that, for all $y \in \mathcal{Y}$,

$$\operatorname{Var}_{\mathbf{p}_z}[W(y \mid X)] \le \mathbb{E}_{\mathbf{p}_z}\left[W(y \mid X)^2\right] \le e^\varepsilon \min_{x \in \mathcal{X}} W(y \mid x)\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)].$$

Hence

$$\int_{\mathcal{Y}} \frac{\operatorname{Var}_{\mathbf{p}_z}[W(y \mid X)]}{\mathbb{E}_{\mathbf{p}_z}[W(y \mid X)]}\,\mathrm{d}\mu \le e^\varepsilon \int_{\mathcal{Y}} \min_{x \in \mathcal{X}} W(y \mid x)\,\mathrm{d}\mu \le e^\varepsilon \cdot \min_{x \in \mathcal{X}} \int_{\mathcal{Y}} W(y \mid x)\,\mathrm{d}\mu = e^\varepsilon.$$

The bound (7) (under Assumption 3) will follow from (5), and the relation between differential privacy and KL divergence. Indeed, the mutual information $I(\mathbf{p}_z; W)$ can be rewritten as the expected (over $X \sim \mathbf{p}_Z$) KL divergence between the distribution $\mathbf{p}^{W,X} := W(\cdot \mid X)$ over $\mathcal{Y}$ induced by the channel W on input X, and the distribution $\mathbf{p}_Z^W := \mathbb{E}_{X' \sim \mathbf{p}_z}[W(\cdot \mid X')]$ over $\mathcal{Y}$ induced by the input distribution $\mathbf{p}_z$ and the channel W:

$$I(\mathbf{p}_z; W) = \mathbb{E}_{X \sim \mathbf{p}_z}\left[\mathrm{D}\left(\mathbf{p}^{W,X} \| \mathbf{p}_z^W\right)\right] = \mathbb{E}_{X \sim \mathbf{p}_z}\left[\mathbb{E}_{Y \sim \mathbf{p}^{W,X}}\left[\ln \frac{W(Y \mid X)}{\mathbb{E}_{X' \sim \mathbf{p}_z}[W(Y \mid X')]}\right]\right];$$

but the ε-LDP condition from Eq. (2) guarantees that the log-likelihood ratio in the inner expectation is (almost surely) at most ε, so that $I(\mathbf{p}_z; W) \le \varepsilon$ for every z and $W \in \mathcal{W}^{\text{priv},\varepsilon}$. This yields (7).

E.4 Proof of Corollary 2

In view of (4), to establish (8), it suffices to show that $\frac{\operatorname{Var}_{\mathbf{p}_z}[W(y|X)]}{\mathbb{E}_{\mathbf{p}_z}[W(y|X)]} \le 1$ for every $y \in \mathcal{Y}$. Since $W(y \mid x) \in (0, 1]$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, so that

$$\operatorname{Var}_{\mathbf{p}_z}[W(y \mid X)] \le \mathbb{E}_{\mathbf{p}_z}\left[W(y \mid X)^2\right] \le \mathbb{E}_{\mathbf{p}_z}[W(y \mid X)].$$

The second bound (under Assumption 3) will follow from (5). Indeed, recalling that the entropy of the output of a channel is bounded below by the mutual information between input and the output, we have $I(\mathbf{p}_z; W) \le H(\mathbf{p}_z^W)$, where $\mathbf{p}_z^W := \mathbb{E}_{\mathbf{p}_z}[W(\cdot \mid X)]$ is the distribution over $\mathcal{Y}$ induced by the input distribution $\mathbf{p}_z$ and the channel W. Using the fact that the entropy of a distribution over $\mathcal{Y}$ is at most $\log |\mathcal{Y}|$ in (5) gives (9).

F Missing proofs in Section 4

F.1 Proof of Lemma 1

Given an (n, γ)-estimator $(\Pi, \hat{\theta})$, define an estimate $\hat{Z}$ for Z as

$$\hat{Z} := \operatorname*{argmin}_{z \in \mathcal{Z}} \left\| \theta_z - \hat{\theta}(Y^n, U) \right\|_p.$$

By the triangle inequality,

$$\left\| \theta_Z - \theta_{\hat{Z}} \right\|_p \le \left\| \theta_Z - \hat{\theta}(Y^n, U) \right\|_p + \left\| \theta_{\hat{Z}} - \hat{\theta}(Y^n, U) \right\|_p \le 2 \left\| \hat{\theta}(Y^n, U) - \theta_Z \right\|_p.$$

Since $(\Pi, \hat{\theta})$ is an (n, γ)-estimator under ℓ_p loss for $\mathcal{P}_\Theta$,

$$\begin{aligned}
\mathbb{E}_Z \left[\mathbb{E}_{\mathbf{p}_Z} \left[\left\| \theta_Z - \theta_{\hat{Z}} \right\|_p^p \right] \right] &\le 2^p \gamma^p \Pr[\, \mathbf{p}_Z \in \mathcal{P}_\Theta \,] + \max_{z \ne z'} \| \theta_z - \theta_{z'} \|_p^p \Pr[\, \mathbf{p}_Z \notin \mathcal{P}_\Theta \,] \\
&\le 2^p \gamma^p + 4^p \gamma^p \frac{1}{\tau} \cdot \frac{\tau}{4} && (35) \\
&\le \frac{3}{4} 4^p \gamma^p, && (36)
\end{aligned}$$

where Eq. (35) follows from Assumption 4 and $\Pr[\, \mathbf{p}_Z \in \mathcal{P}_\Theta \,] \ge 1 - \tau/4$. Next, for $p \in [1, \infty)$, by Assumption 4, $\left\| \theta_Z - \theta_{\hat{Z}} \right\|_p^p \ge \frac{4^p \gamma^p}{\tau k} \sum_{i=1}^k \mathbb{1}\left\{ Z_i \ne \hat{Z}_i \right\}$. Combining with Eq. (36) this shows that $\frac{1}{\tau k} \sum_{i=1}^k \Pr\left[\, Z_i \ne \hat{Z}_i \,\right] \le \frac{3}{4}$.

Furthermore, since the Markov relation $Z_i - (Y^n, U) - \hat{Z}_i$ holds for all i, we can lower bound $\Pr\left[\, Z_i \ne \hat{Z}_i \,\right]$ using the standard relation between total variation distance and hypothesis testing as follows, using that $\tau \le 1/2$ in the second inequality:

$$\begin{aligned}
\Pr\left[\, Z_i \ne \hat{Z}_i \,\right] &\ge \tau \Pr\left[\, \hat{Z}_i = -1 \,\middle|\, Z_i = 1 \,\right] + (1 - \tau) \Pr\left[\, \hat{Z}_i = 1 \,\middle|\, Z_i = -1 \,\right] \\
&\ge \tau \left(\Pr\left[\, \hat{Z}_i = -1 \,\middle|\, Z_i = 1 \,\right] + \Pr\left[\, \hat{Z}_i = 1 \,\middle|\, Z_i = -1 \,\right] \right) \\
&\ge \tau \left(1 - \mathrm{d}_{\mathrm{TV}}\left(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n} \right) \right).
\end{aligned}$$

Summing over $1 \le i \le k$ and combining it with the previous bound, we obtain

$$\frac{3}{4} \ge \frac{1}{\tau k} \sum_{i=1}^k \Pr\left[\, Z_i \ne \hat{Z}_i \,\right] \ge 1 - \frac{1}{k} \sum_{i=1}^k \mathrm{d}_{\mathrm{TV}}\left(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n} \right)$$

and reorganizing proves the result.

G Missing statements and proofs in Section 5

G.1 Proof of Theorem 3

Fix $p \in [1, \infty)$. Let $k = d$, $\mathcal{Z} = \{-1, +1\}^d$, and $\tau = \frac{s}{2d}$; and suppose that, for some $\gamma \in (0, 1/8]$, there exists an (n, γ)-estimator for $\mathcal{B}_{d,s}$ under ℓ_p loss. We fix a parameter $\gamma \in (0, 1/2]$, which will be chosen as a function of γ, d, p later. Consider the set of 2^d product Bernoulli distributions $\{\mathbf{p}_z\}_{z \in \mathcal{Z}}$, where $\mu(\mathbf{p}_z) = \mu_z := \frac{1}{2} \gamma (z + \mathbf{1}_d)$ (so the sparsity of the mean vector is equal to the number of positive coordinates of z). We have, for $z \in \mathcal{Z}$,

$$\mathbf{p}_z(x) = \frac{1}{2^d} \prod_{i=1}^d \left(1 + \frac{1}{2} \gamma (z_i + 1) x_i \right), \qquad x \in \mathcal{X}.$$

It follows for $z \in \mathcal{Z}$ and $i \in [d]$ that

$$\mathbf{p}_{z^{\oplus i}}(x) = \frac{1 + \frac{1}{2}\gamma(1 - z_i)x_i}{1 + \frac{1}{2}\gamma(1 + z_i)x_i}\mathbf{p}_z(x) = \left(1 - \gamma\frac{z_i x_i}{1 + \frac{1}{2}\gamma(1 + z_i)x_i}\right)\mathbf{p}_z(x) = (1 + \phi_{z,i}(x))\mathbf{p}_z(x) \tag{37}$$

where $\phi_{z,i}(x) := -\frac{\gamma z_i x_i}{1+\frac{1}{2}\gamma(1+z_i)x_i}$. We can verify that, for $i \neq j$,

$$\mathbb{E}_{\mathbf{p}_z}[\phi_{z,i}(X)] = 0, \quad \mathbb{E}_{\mathbf{p}_z}\left[\phi_{z,i}(X)^2\right] = \frac{\gamma^2}{1 - \frac{1}{2}\gamma^2(1 + z_i)}, \text{ and } \mathbb{E}_{\mathbf{p}_z}[\phi_{z,i}(X)\phi_{z,j}(X)] = 0,$$

so that Assumptions 1 and 2 are satisfied for $\alpha^2 := 2\gamma^2$. Moreover, using, *e.g.*, Hoeffding's lemma (*cf.* [9]), for $\gamma < 1$, the random vector $\phi_z(X) = (\phi_{z,i}(X))_{i\in[d]}$ is $\frac{\gamma^2}{(1-\gamma^2)^2}$-subgaussian. Thus, Assumption 3 holds as well, and we can invoke both parts of Theorem 2.

Let $\|z\|_+ := |\{i \in [d] \mid z_i = 1\}|$, so that $\|\mu_z\|_0 = \sum_{i=1}^d \frac{1}{2}(1 + z_i) = \|z\|_+$. The next claim, which follows from standard bounds for binomial random variables, states that when $Z \sim \mathrm{Rad}(\tau)^{\otimes d}$, μ_Z is s-sparse with high probability.

Fact 2. *Let $Z \sim \mathrm{Rad}(\tau)^{\otimes d}$, where $\tau d \geq 4\log d$. Then* $\Pr\left[\, \|Z\|_+ \leq 2\tau d \,\right] \geq 1 - \tau/4$.

Hence the construction satisfies $\Pr_Z[\, \mathbf{p}_Z \in \mathcal{B}_{d,s}\,] \leq 1 - \tau/4$, as required in Lemma 1.

We now choose $\gamma = \gamma(p) := \frac{4\gamma}{(s/2)^{1/p}} \in (0, 1/2]$, which implies that Assumption 4 holds since

$$\ell_p(\mu(\mathbf{p}_z), \mu(\mathbf{p}_{z'})) = \gamma\, \mathrm{d_{Ham}}(z, z')^{1/p} = 4\gamma\left(\frac{\mathrm{d_{Ham}}(z, z')}{\tau d}\right)^{1/p}.$$

Therefore, we can apply Lemma 1 as well. For $\mathcal{W}^{\mathrm{priv},\varepsilon}$, we prove the two parts of the lower bound separately, depending on whether $\varepsilon \leq 1$. First, upon combining the bounds obtained by Corollary 1 and Lemma 1 (specifically, for the former, (6)), we get

$$d \leq 112n\alpha^2(e^\varepsilon - 1)^2,$$

whereby, upon recalling that $\alpha^2 = 2\gamma^2$, and using the value of $\gamma = \gamma(p)$ above, it follows that

$$\frac{1}{3584} \cdot \frac{d(s/2)^{\frac{2}{p}}}{n(e^\varepsilon - 1)^2} \leq \gamma^2.$$

Thus, $\mathcal{E}_p(\mathcal{B}_{d,s}, \mathcal{W}^{\mathrm{priv},\varepsilon}, n) = \Omega\left(\sqrt{\frac{ds^{2/p}}{n\varepsilon^2}}\right)$ for $\varepsilon \in (0, 1]$. For the second part of the bound, which dominates for $\varepsilon > 1$, observe that Assumption 3 holds with $\sigma^2 := \frac{\gamma^2}{(1-\gamma^2)^2} \leq 2\gamma^2$; allowing us to apply the second part of Corollary 1, (7), which as before combined with Lemma 1 yields

$$d \leq 224n\sigma^2\varepsilon \leq 448n\gamma^2\varepsilon,$$

and again from the setting of γ we get $\mathcal{E}_p(\mathcal{B}_{d,s}, \mathcal{W}^{\mathrm{priv},\varepsilon}, n) = \Omega\left(\sqrt{\frac{ds^{2/p}}{n\varepsilon}}\right)$.

Similarly, for $\mathcal{W}^{\mathrm{comm},\ell}$, again since Assumption 3 holds with $\sigma^2 \leq 2\gamma^2$, upon combining the bounds obtained by Corollary 2 and Lemma 1, we get

$$\frac{ds^{\frac{2}{p}}}{28672n\ell} \leq \gamma^2,$$

which gives $\mathcal{E}_p(\mathcal{B}_{d,s}, \mathcal{W}^{\mathrm{comm},\ell}, n) = \Omega\big(\sqrt{\frac{ds^{2/p}}{n\ell}} \wedge 1\big)$. Finally, note that for $\ell \geq d$, the lower bound follows from the minimax rate in the unconstrained setting, which can be seen to be $\Omega\big(\sqrt{s^{2/p}\log(2d/s)/n}\big)$ [28, 30]. This completes the proof.

This handles the case $p \in [1, \infty)$. For $p = \infty$, the lower bounds immediately follow from plugging $p = \log s$ in the previous expressions, as discussed in Footnote 3.

G.2 Detailed results for Gaussian family

Similar to the previous section, we denote the mean by μ instead of θ, denote the estimator by $\hat{\mu}$, and consider the minimax error rate $\mathcal{E}_p(\mathcal{G}_{d,s}, \mathcal{W}, n)$ of mean estimation for $\mathcal{P}_\Theta = \mathcal{G}_{d,s}$ using $\mathcal{W}$ under ℓ_p loss.

We derive a lower bound for $\mathcal{E}_p(\mathcal{G}_{d,s}, \mathcal{W}, n)$ under local privacy (captured by $\mathcal{W} = \mathcal{W}^{\text{priv},\varepsilon}$) and communication (captured by $\mathcal{W} = \mathcal{W}^{\text{comm},\ell}$) constraints.[9] Recall that for product Bernoulli mean estimation we had optimal bounds for both privacy and communication constraints for all finite p. For Gaussians, we will obtain tight bounds for privacy constraints for $\varepsilon \in (0, 1]$. However, for communication constraints and privacy constraints when $\varepsilon \geq 1$, our bounds for Gaussian distributions are tight only in specific regimes of n up to logarithmic factors. We state our general result and provide some remarks before providing the proofs.

We defer the estimation schemes and their analysis (*i.e.*, upper bounds) to the Supplement (Appendix C.2); they follow from a simple reduction from the Gaussian estimation problem to the product Bernoulli one, which enables us to invoke the protocols for the latter task in both the communication-constrained and locally private settings.

Theorem 5. *Fix $p \in [1, \infty)$. For $4 \log d \leq s \leq d$, under LDP constraints, when $\varepsilon \in (0, 1]$,*

$$\sqrt{\frac{ds^{2/p}}{n\varepsilon^2}} \wedge 1 \lesssim \mathcal{E}_p(\mathcal{G}_{d,s}, \mathcal{W}^{\text{priv},\varepsilon}, n) \lesssim \sqrt{\frac{ds^{2/p}}{n\varepsilon^2}} \tag{38}$$

and when $\varepsilon > 1$,

$$\sqrt{\frac{ds^{2/p}}{n\varepsilon \log{(nd)}} \vee \frac{s^{2/p} \log \frac{2d}{s}}{n}} \wedge 1 \lesssim \mathcal{E}_p(\mathcal{G}_{d,s}, \mathcal{W}^{\text{priv},\varepsilon}, n) \lesssim \sqrt{\frac{ds^{2/p}}{n\varepsilon} \vee \frac{s^{2/p} \log \frac{2d}{s}}{n}} \tag{39}$$

Under communication constraints,

$$\sqrt{\frac{ds^{2/p}}{n\ell \log(dn)} \vee \frac{s^{2/p} \log \frac{2d}{s}}{n}} \wedge 1 \lesssim \mathcal{E}_p(\mathcal{G}_{d,s}, \mathcal{W}^{\text{comm},\ell}, n) \lesssim \sqrt{\frac{ds^{2/p}}{n\ell} \vee \frac{s^{2/p} \log \frac{2d}{s}}{n}} \tag{40}$$

For $p = \infty$, we have the upper bounds

$$\mathcal{E}_\infty(\mathcal{G}_{d,s}, \mathcal{W}^{\text{priv},\varepsilon}, n) = O\left(\sqrt{\frac{d \log s}{n\varepsilon^2}}\right) \quad \textit{and} \quad \mathcal{E}_\infty(\mathcal{G}_{d,s}, \mathcal{W}^{\text{comm},\ell}, n) = O\left(\sqrt{\frac{d \log s}{n\ell} \vee \frac{\log d}{n}}\right),$$

while the lower bounds given in Eqs. (38), (39), *and* (40) *hold for $p = \infty$, too.*[10]

We emphasize that, as discussed in Sections 1.1 and 1.2, to the best of our knowledge Theorem 5 provides the first lower bounds for interactive Gaussian mean estimation under communication and privacy constraints.

Proof of Theorem 5. Let φ denote the probability density function of the standard Gaussian distribution $\mathcal{G}(\mathbf{0}, \mathbb{I})$. Fix $p \in [1, \infty)$. Let $k = d$, $\mathcal{Z} = \{-1, +1\}^d$, and $\tau = \frac{s}{2d}$; and suppose that, for some $\gamma \in (0, 1/8]$, there exists an (n, γ)-estimator for $\mathcal{G}_{d,s}$ under ℓ_p loss. We fix a parameter $\gamma := \gamma(p) := \frac{4\gamma}{(s/2)^{1/p}} \in (0, 1/2]$, and consider the set of distributions $\{\mathbf{p}_z\}_{z \in \mathcal{Z}}$ of all 2^d spherical Gaussian distributions with mean $\mu_z := \gamma(z + \mathbf{1}_d)$, where $z \in \mathcal{Z}$. Again, note that $\|\mu_z\|_0 = \sum_{i=1}^d \mathbb{1}\{z_i = 1\} = \|z\|_+$, and Fact 2 applies here too. Then by the definition of Gaussian density, for $z \in \mathcal{Z}$,

$$\mathbf{p}_z(x) = e^{-\gamma^2 \|\mu_z\|_2^2/2} \cdot e^{\gamma \langle x, z + \mathbf{1}_d \rangle} \cdot \varphi(x). \tag{41}$$

Therefore, for $z \in \mathcal{Z}$ and $i \in [d]$, we have

$$\mathbf{p}_{z^{\oplus i}}(x) = e^{-2\gamma x_i z_i} e^{2\gamma^2 z_i} \cdot \mathbf{p}_z(x) = (1 + \phi_{z,i}(x)) \cdot \mathbf{p}_z(x), \tag{42}$$

[9] As in the Bernoulli case, we here focus for simplicity on the case where the communication (resp., privacy) parameters are the same for all players, but our lower bounds easily extend.

[10] That is, the upper and lower bounds only differ by a $\log s$ factor for $p = \infty$ in the privacy case.

where $\phi_{z,i}(x) := 1 - e^{-2\gamma x_i z_i} e^{2\gamma^2 z_i}$. By using the Gaussian moment-generating function, for $i \neq j$,

$$\mathbb{E}_{\mathbf{p}_z}[\phi_{z,i}(X)] = 0, \quad \mathbb{E}_{\mathbf{p}_z}\left[\phi_{z,i}(X)^2\right] = e^{4\gamma^2} - 1, \text{ and } \mathbb{E}_{\mathbf{p}_z}[\phi_{z,i}(X)\phi_{z,j}(X)] = 0,$$

so that Assumptions 1 and 2 are satisfied for $\alpha^2 := e^{4\gamma^2} - 1$. By our choice of γ and the assumption on γ, one can check that Assumption 4 holds:

$$\ell_p(\mu(\mathbf{p}_z), \mu(\mathbf{p}_{z'})) = 4\gamma \left(\frac{\mathrm{d_{Ham}}(z, z')}{\tau d} \right)^{1/p}.$$

Moreover, similar to the product of Bernoulli case, using Fact 2, we can show that $\Pr_Z[\, \mathbf{p}_Z \in \mathcal{G}_{d,s} \,] \leq 1 - \tau/4$. This allows us to apply Lemma 1.

G.2.1 Privacy constraints for $\varepsilon \in (0, 1)$

For $\mathcal{W}^{\mathrm{priv},\varepsilon}$, upon combining the bounds obtained by Corollary 1 and Lemma 1, we get

$$d \leq 112 n \alpha^2 (e^{\varepsilon} - 1)^2,$$

whereby, upon noting that $\alpha^2 = e^{4\gamma^2} - 1 \leq 8\gamma^2$ holds since $\gamma \leq 1/2$, and using the value of $\gamma = \gamma(p)$ above, it follows that

$$\gamma^2 \geq \frac{d (s/2)^{\frac{2}{p}}}{14336 \cdot n (e^{\varepsilon} - 1)^2}.$$

Thus, $\mathcal{E}_p(\mathcal{G}_{d,s}, \mathcal{W}^{\mathrm{priv},\varepsilon}, n) = \Omega\left(\sqrt{\frac{d s^{2/p}}{n \varepsilon^2}} \wedge 1 \right)$. This establishes the lower bounds for $\mathcal{W}^{\mathrm{priv},\varepsilon}$. (Recall that the bound for $p = \infty$ then follows from setting $p = \log d$.)

G.2.2 Communication constraints, and privacy constraints for $\varepsilon \geq 1$

For these cases, to prove a lower bound with the desired dependence on ε or ℓ, we will need to use the tighter bounds in Corollaries 1 and 2 which hold only under Assumption 3. This, however, leads to an issue: the random vector $\phi_z(X) = (\phi_{z,i}(X))_{i \in [d]}$ is not subgaussian, due to the one-sided exponential growth, and therefore Assumption 3 does not hold.

To overcome this and still obtain a linear dependence on ℓ (or ε) (instead of the suboptimal 2^{ℓ} (or e^{ε})), we will consider instead the class of "truncated" Gaussian distributions, whose corresponding ϕ functions are subgaussian; and argue that these truncated distributions are close enough to the original Gaussian distributions such a lower bound in the truncated case implies one in the original Gaussian case.

In particular, we consider the following collection of truncated Gaussian distributions. For $z \in \mathcal{Z}$, let $\mathbf{p}_z$ be the density function of a spherical Gaussian distribution with mean μ_z as defined in Eq. (41). For a truncation bound B, let $\mathbf{p}_{z,B}$ be the distribution of $X \sim \mathbf{p}_z$ conditioned on the event that $\|X\|_\infty \leq B$. That is, we have, for $x \in \mathbb{R}^d$,

$$\mathbf{p}_{z,B}(x) = C_z \mathbf{p}_z(x) \mathbb{1}\{\|X\|_\infty \leq B\},$$

where $C_z = 1/\Pr_{X \sim \mathbf{p}_z}[\, \|X\|_\infty \leq B \,]$. Then the following bound follows from standard Gaussian concentration bound on each dimension and a union bound over all dimensions.

Fact 3. *Setting $B := 4\sqrt{\ln(dn)}$, we have, for every $z \in \mathcal{Z}$, $\mathrm{d_{TV}}(\mathbf{p}_{z,B}, \mathbf{p}_z) \leq \frac{1}{d^7 n^8}$.*

Let $\mathbf{p}_{z,B}^{Y^n}$ be the distribution of the messages obtained by executing the protocol when each user gets a sample from $\mathbf{p}_{z,B}$ and let the corresponding mixtures be denoted by $\mathbf{p}_{+i,B}^{Y^n}$ and $\mathbf{p}_{-i,B}^{Y^n}$. Then we have

$$\begin{aligned}
\mathrm{d_{TV}}\left(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{-i}^{Y^n}\right) &\leq \mathrm{d_{TV}}\left(\mathbf{p}_{+i,B}^{Y^n}, \mathbf{p}_{-i,B}^{Y^n}\right) + \mathrm{d_{TV}}\left(\mathbf{p}_{+i}^{Y^n}, \mathbf{p}_{+i,B}^{Y^n}\right) + \mathrm{d_{TV}}\left(\mathbf{p}_{-i,B}^{Y^n}, \mathbf{p}_{-i}^{Y^n}\right) \\
&\leq \mathrm{d_{TV}}\left(\mathbf{p}_{+i,B}^{Y^n}, \mathbf{p}_{-i,B}^{Y^n}\right) + \max_z \left\{ \mathrm{d_{TV}}\left(\mathbf{p}_{z}^{Y^n}, \mathbf{p}_{z,B}^{Y^n}\right) + \mathrm{d_{TV}}\left(\mathbf{p}_{z,B}^{Y^n}, \mathbf{p}_{z}^{Y^n}\right) \right\} \\
&\leq \mathrm{d_{TV}}\left(\mathbf{p}_{+i,B}^{Y^n}, \mathbf{p}_{-i,B}^{Y^n}\right) + 2 \max_z \mathrm{d_{TV}}\left(\mathbf{p}_{z,B}^{\otimes n}, \mathbf{p}_{z}^{\otimes n}\right) \\
&\leq \mathrm{d_{TV}}\left(\mathbf{p}_{+i,B}^{Y^n}, \mathbf{p}_{-i,B}^{Y^n}\right) + 2n \max_z \mathrm{d_{TV}}(\mathbf{p}_{z,B}, \mathbf{p}_z) \\
&\leq \mathrm{d_{TV}}\left(\mathbf{p}_{+i,B}^{Y^n}, \mathbf{p}_{-i,B}^{Y^n}\right) + \frac{2}{d^7 n^7}.
\end{aligned}$$

The third inequality follows from data processing inequality and the fourth inequality follows from subadditivity of TV distance.

Combining this with Lemma 1, for any protocol that correctly learns the Gaussian family, we must have

$$\frac{1}{d}\sum_{i=1}^{d} \mathrm{d_{TV}}\left(\mathbf{p}_{+i,B}^{Y^n}, \mathbf{p}_{-i,B}^{Y^n}\right) \geq \frac{1}{8}. \tag{43}$$

Next we show that the ϕ functions corresponding to $\mathbf{p}_{z,B}$'s are subgaussian and establish the corresponding upper bounds on the average information bound above. Note that

$$\phi_{z,i}^{B}(x) := \frac{\mathbf{p}_{z^{\oplus i}}^{B}(x)}{\mathbf{p}_{z}^{B}(x)} - 1 = \frac{C_{z^{\oplus i}}}{C_z} e^{-2\gamma x_i z_i} e^{2\gamma^2 z_i} \mathbb{1}\{\|x\|_\infty \leq B\} - 1 \tag{44}$$

By the inequality $|ab-1| \leq |a|\cdot|b-1| + |a-1|$, we have have, for all $z \in \mathcal{Z}$,

$$\left|\frac{C_{z^{\oplus i}}}{C_z} - 1\right| \leq \frac{1}{C_z}|C_{z^{\oplus i}} - 1| + \left|\frac{1}{C_z} - 1\right| \leq \left|\frac{1}{\Pr_{X\sim\mathbf{p}_{z^{\oplus i}}}[\,\|X\|_\infty \leq B\,]} - 1\right| + \left|\Pr_{X\sim\mathbf{p}_z}[\,\|X\|_\infty \leq B\,] - 1\right|$$
$$\leq \frac{10}{d^7 n^7}.$$

Moreover, for all $z \in \mathcal{Z}$, for $\gamma \leq \frac{1}{3B}$,

$$\left|e^{-2\gamma x_i z_i} e^{2\gamma^2 z_i} \mathbb{1}\{\|x\|_\infty \leq B\} - 1\right| \leq \left|e^{2\gamma^2 + 2\gamma B} - 1\right| \leq \left|e^{3\gamma B} - 1\right| \leq 6\gamma B. \tag{45}$$

Hence, applying the inequality $|ab-1| \leq |a|\cdot|b-1| + |a-1|$ again on Eq. (44), we have for $\gamma \leq \frac{1}{3B}$,

$$|\phi_{z,i}^{B}(x)| \leq 12\gamma B + \frac{10}{d^7 n^7}.$$

Thus, we get that for all $z \in \mathcal{Z}, i \in [d]$, $\phi_{z,i}^{B}$ is subgaussian with proxy $\sigma_B = 12\gamma B + \frac{10}{d^7 n^7}$.

Under communication constraints, applying Corollary 2, we get

$$\left(\frac{1}{d}\sum_{i=1}^{d} \mathrm{d_{TV}}\left(\mathbf{p}_{+i,B}^{Y^n}, \mathbf{p}_{-i,B}^{Y^n}\right)\right)^2 \leq \frac{14}{d}\sigma_B^2 n\ell.$$

To conclude, we observe that by plugging our setting of $\gamma = \gamma(p)$ in the above inequality, we must have

$$\gamma^2 \geq \frac{d(s/2)^{\frac{2}{p}}}{14336 \cdot n \cdot B^2 \ell}$$

in order to satisfy Eq. (43), hence proving the desired lower bound. The lower bound for LDP with $\varepsilon > 1$ follows similarly by applying Corollary 1. □

G.3 Detailed results for discrete family

We derive a lower bound for $\mathcal{E}_p(\Delta_d, \mathcal{W}, n)$, the minimax rate for discrete density estimation, under local privacy and communication constraints.

Theorem 6. *Fix $p \in [1,\infty)$. For $\varepsilon > 0$, and $\ell \geq 1$, we have*

$$\mathcal{E}_p(\Delta_d, \mathcal{W}^{\text{priv},\varepsilon}, n) \gtrsim \sqrt{\frac{d^{2/p}}{n((e^\varepsilon-1)^2 \wedge e^\varepsilon)} \wedge \left(\frac{1}{n((e^\varepsilon-1)^2 \wedge e^\varepsilon)}\right)^{\frac{p-1}{p}}} \wedge 1 \tag{46}$$

and

$$\mathcal{E}_p(\Delta_d, \mathcal{W}^{\text{comm},\ell}, n) \gtrsim \sqrt{\frac{d^{2/p}}{n2^\ell} \wedge \left(\frac{1}{n2^\ell}\right)^{\frac{p-1}{p}}} \wedge 1\,. \tag{47}$$

In particular, for $n\big((e^\varepsilon-1)^2 \wedge e^\varepsilon\big) \geq d^2$ and $n(2^\ell \wedge d) \geq d^2$, the first term of the corresponding lower bounds dominates. Before turning to the proof of this theorem, we note that Corollary 3 and Corollary 4 are direct corollaries of the theorem.

We now establish Theorem 6.

Proof of Theorem 6. Fix $p \in [1, \infty)$, and suppose that, for some $\gamma \in (0, 1/16]$, there exists an (n, γ)-estimator for Δ_d under ℓ_p loss. Set

$$D := d \wedge \left\lfloor \left(\frac{1}{16\gamma}\right)^{\frac{p}{p-1}} \right\rfloor$$

and assume, without loss of generality, that D is even. By definition, we then have $\gamma \in (0, 1/(16D^{1-1/p})]$ and $D \le d$; we can therefore restrict ourselves to the first D elements of the domain, embedding Δ_D into Δ_d, to prove our lower bound.

Let $k = \frac{D}{2}$, $\mathcal{Z} = \{-1, +1\}^{D/2}$, and $\tau = \frac{1}{2}$; and suppose that, for some $\gamma \in (0, 1/(16D^{1-1/p})]$, there exists an (n, γ)-estimator for Δ_D under ℓ_p loss. (We will use the fact that $\gamma \le 1/(16D^{1-1/p})$ for Eq. (49) to be a valid distribution with positive mass, as we will need $|\gamma| \le \frac{1}{D}$; and to bound α^2 later on, as we will require $|\gamma| \le \frac{1}{2D}$.) Define $\gamma = \gamma(p)$ as

$$\gamma(p) := \frac{4 \cdot 2^{1/p}\gamma}{D^{1/p}}, \tag{48}$$

which implies $\gamma \in [0, 1/(2D)]$. Consider the set of D-ary distributions $\mathcal{P}^{\gamma}_{\text{Discrete}} = \{\mathbf{p}_z\}_{z \in \mathcal{Z}}$ defined as follows. For $z \in \mathcal{Z}$, and $x \in \mathcal{X} = [D]$

$$\mathbf{p}_z(x) = \begin{cases} \frac{1}{D} + \gamma z_i, & \text{if } x = 2i, \\ \frac{1}{D} - \gamma z_i, & \text{if } x = 2i - 1. \end{cases} \tag{49}$$

For $z \in \mathcal{Z}$ and $i \in [D/2]$, we have

$$\begin{aligned} \mathbf{p}_{z^{\oplus i}}(x) &= \left(1 - \frac{2D\gamma z_i}{1 + D\gamma z_i}\mathbb{1}\{x = 2i\} + \frac{2D\gamma z_i}{1 - D\gamma z_i}\mathbb{1}\{x = 2i - 1\}\right)\mathbf{p}_z(x) \\ &= (1 + \phi_{z,i}(x))\mathbf{p}_z(x), \end{aligned} \tag{50}$$

where

$$\phi_{z,i}(x) := z_i \cdot \frac{2D\gamma}{1 - D^2\gamma^2}((1 + D\gamma z_i)\mathbb{1}\{x = 2i - 1\} - (1 - D\gamma z_i)\mathbb{1}\{x = 2i\}).$$

Once again, we can verify that for $i \neq j$

$$\mathbb{E}_{\mathbf{p}_z}[\phi_{z,i}(X)] = 0, \quad \mathbb{E}_{\mathbf{p}_z}\left[\phi_{z,i}(X)^2\right] = \frac{8\gamma^2 D}{1 - \gamma^2 D^2}, \text{ and } \mathbb{E}_{\mathbf{p}_z}[\phi_{z,i}(X)\phi_{z,j}(X)] = 0,$$

so that Assumptions 1 and 2 are satisfied for $\alpha^2 := 16\gamma^2 D$ (using that $D\gamma \le 1/2$ to simplify the bound).[11] Thus, we can invoke the first part of Theorem 2. Note that Assumption 4 holds, since $\ell_p(\mathbf{p}_z, \mathbf{p}_{z'}) = \gamma \, \mathrm{d}_{\text{Ham}}(z, z')^{1/p} = 4\gamma\left(\frac{\mathrm{d}_{\text{Ham}}(z,z')}{\tau D}\right)^{1/p}$. Therefore, we can apply Lemma 1 as well.

For $\mathcal{W}^{\text{priv},\varepsilon}$, by combining the bounds obtained by Corollary 1 and Lemma 1, we get

$$D \le 56n\alpha^2\left((e^\varepsilon - 1)^2 \wedge e^\varepsilon\right),$$

whereby, upon recalling the value of α^2 and using the setting of $\gamma = \gamma(p)$ from Eq. (48), it follows that

$$\gamma^2 \ge \frac{D^{\frac{2}{p}}}{7168 \cdot 2^{2/p} \cdot n((e^\varepsilon - 1)^2 \wedge e^\varepsilon)} \asymp \frac{d^{2/p} \wedge \gamma^{-2/(p-1)}}{n((e^\varepsilon - 1)^2 \wedge e^\varepsilon)}.$$

Thus we obtain the bound Eq. (46) as claimed.

Similarly, for $\mathcal{W}^{\text{comm},\ell}$, upon combining the bounds obtained by Corollary 2 and Lemma 1 and recalling that $|\mathcal{Y}| = 2^\ell$, we get

$$\gamma^2 \ge \frac{D^{\frac{2}{p}}}{7168 \cdot 2^{2/p} \cdot n2^\ell},$$

which gives $\mathcal{E}_p(\Delta_D, \mathcal{W}^{\text{comm},\ell}, n) = \Omega\left(\sqrt{\frac{d^{2/p}}{n2^\ell} \wedge \left(\frac{1}{n2^\ell}\right)^{\frac{p-1}{p}}}\right)$,[12] concluding the proof. □

[11] It is worth noting that Assumption 3 will not hold for any useful choice of the subgaussianity parameter.

[12] Finally, note that we could replace the quantity 2^ℓ above by $2^\ell \wedge d$, or even $2^\ell \wedge D$, as for $2^\ell \ge D$ there is no additional information any player can send beyond the first $\log_2 D$ bits, which encode their full observation. However, this small improvement would lead to more cumbersome expressions, and not make any difference for the main case of interest, $p = 1$.