On Learning Parametric Distributions from Quantized Samples

Septimia Sarbu *, Abdellatif Zaidi †*
*Huawei Technologies France, Mathematical and Algorithmic Sciences Lab,
Paris Research Center, Boulogne-Billancourt, 92100, France
†Université Paris-Est, Champs-sur-Marne, 77454, France
{septimia.sarbu@huawei.com, abdellatif.zaidi@u-pem.fr}

Abstract—We consider the problem of learning parametric distributions from their quantized samples in a network. Specifically, $n$ agents or sensors observe independent samples of an unknown parametric distribution; and each of them uses $k$ bits to describe its observed sample to a central processor whose goal is to estimate the unknown distribution. First, we establish a generalization of the well-known van Trees inequality to general $L_p$-norms, with $p > 1$, in terms of Generalized Fisher information. Then, we develop minimax lower bounds on the estimation error for two losses: general $L_p$-norms and the related Wasserstein loss from optimal transport.

I. INTRODUCTION AND PROBLEM FORMULATION

Consider the multiterminal detection system shown in Figure 1. In this problem a memoryless vector source $X$ has joint distribution $f(x|\theta)$ that depends on an unknown (vector) parameter $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$, with $d \geq 1$. A number of agents or sensors, say $n$, observe each one independent sample of $X$; and each of them uses $k \geq 1$ bits to describe its sample to a fusion center whose goal is to find a distribution $\hat{f}$ that approximates the unknown (parametric) distribution $f(x|\theta)$ in a suitable sense. How well can $f(x|\theta)$ be approximated from the quantized samples? This question has so far been resolved (partially) only for few special cases, among which the $L_2$ loss [1, 2]. Worse, even in the extreme case in which $k$ is large (quantized samples) little is known about this problem for general loss measures [3].

![Fig. 1. Distribution estimation from quantized samples.](image)

In this paper we study an instance of this problem under general $L_p$-norms, where $p \in \mathbb{R}$ with $p > 1$, as well as the related Wasserstein distance of order $p$. We recall that for given distributions $P$ and $Q$, the $p$-Wasserstein distance between $P$ and $Q$ is defined as [6]

$$W_p(P, Q) = \inf_{\nu \in \Pi(P, Q)} \left( \mathbb{E}(Z, Y) \sim \nu [d^p(Z, Y)] \right)^{\frac{1}{p}}$$

(1)

where the random variables $Z \in \mathcal{Z}$ and $Y \in \mathcal{Y}$ have distributions $P$ and $Q$ respectively, i.e., $Z \sim P$ and $Y \sim Q$; the set $\Pi(P, Q)$ designates the set of measures $\nu$ on $\mathcal{Z} \times \mathcal{Y}$ (called couplings) whose $Z$-marginal and $Y$-marginal coincide with $P$ and $Q$ respectively; and $d : \mathcal{Z} \times \mathcal{Y} \rightarrow \mathbb{R}_+$ is a given distance measure. Specifically, let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} f(x|\theta)$

(2)

where $\theta \in \Theta \subseteq \mathbb{R}^d$. Agent $i$, $i = 1, \ldots, n$, observes the sample $X_i$ and sends a $k$-bit string $M_i$ to the fusion center. We assume that the agents process their observations and communicate with the fusion center simultaneously and independently of each other. The fusion center uses the tuple $M^{(n)} = (M_1, \ldots, M_n)$ to find an estimate $\hat{\theta} : = \hat{\theta}(M^{(n)})$ of the unknown parameter $\theta$; and then approximates the unknown source distribution as $\hat{f}(x|\theta)$.

Our goal is to design the estimate $\hat{\theta}$ so as to minimize the worst case power-$p$ Wasserstein risk, i.e.,

to characterize

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E} \left[ W_p^p (f(x|\hat{\theta}), f(x|\theta)) \right].$$

(3)

When the underlying distance in the Wasserstein risk [3] is based on the $L_p$-norm, it is instrumental to study the following related parameter estimation problem under the $L_p$-norm,

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E} \left[ \| \hat{\theta} - \theta \|_p^p \right].$$

(4)

where $\| \cdot \|_p$ designates the $L_p$ norm.

The main contributions of this paper are as follows. First, we establish a generalization of the well known van Trees inequality [7, p. 72], which is a Bayesian analog of the information inequality, to $L_p$-norms with $p > 1$, in terms of generalized Fisher information of order $p$ [14]. This result, which holds under some mild conditions (see Section I) that are assumed to hold throughout, may be of independent interest in its own right. In particular, its proof is more direct than the traditional methods of Assouad, Fano, or Le Cam [17]. Then, we develop lower bounds on the losses [3] and [4] in terms of the order $p$, the number of samples $n$, the number of quantization bits $k$ and the parameter space $d$. Some of our results generalize those of [3], which are established therein for the $L_2$ loss, to the case of $L_p$ loss for arbitrary $p > 1$. Particularly interesting in these bounds is that, for some example source classes that we study, they decrease with the number of samples at least as $1/(n^{\frac{p}{d}})$; and with $k$ at least as $1/(k^{\frac{p}{d}})$ for some suitable value $r > 0$. Key to the proofs of the results of this paper are some judicious applications of inequalities such as H"older inequality and the Marcinkiewicz-Zygmund inequality [18].

A. Related Works

The problem of statistical estimation in distributed settings has attracted increasing interest in recent years, in part motivated by...
learning applications at the wireless Edge. Most relevant to this paper are the works [2], [3]. In particular, the parameter estimation problem [4] is studied in [3] for the case \( p = 2 \), i.e., the squared \( L_2 \) loss. Specifically, in [2] the authors build upon [3] to derive lower bounds on the risk [4] that account (partially) for the loss of Fisher information (relative to the unquantized setting [3]) that is caused by quantization in the case \( p = 2 \). In doing so, they use the standard van Trees inequality which is a Bayesian version of the well known Cramér-Rao inequality for the Euclidean norm \( L_2 \). In this paper, for the study of the problem [4] for general \( p > 1 \), after generalizing the usual van Trees inequality to general \( L_p \)-norms, essentially we follow the approach of [2]. For non-parametric models of densities over \([0, 1]\) that are Holder continuous of smoothness \( s \in (0, 1] \), [4] provides upper and lower bounds on the worst case error under the \( L_1 \) norm. For more on this and other related works, the reader may refer to [2]–[4] as well as the references mentioned therein. For related works on Wasserstein loss based learning, see, e.g., [9]–[13].

II. FORMAL PROBLEM FORMULATION AND DEFINITIONS

Consider the model shown in Figure 1. Here, there are \( n \) sensors which observe each one sample of a memoryless vector source \( X \). We assume that the underlying distribution or density of \( X \) is parametrized by an unknown vector parameter \( \theta = (\theta_1, \ldots, \theta_d) \in \Theta \subseteq \mathbb{R}^d \) of dimension \( d \geq 1 \); and we write \( f(x) := f(x|\theta) \) for \( \theta \in \Theta \). The samples \( X_1, X_2, \ldots, X_n \) are all independent; and they are processed independently by the sensors. Sensor \( i, i = 1, \ldots, n \), encodes its sample \( X_i \) into a \( k \)-bit string \( M_i \in [1, 2^k] \). A (possibly stochastic) \( k \)-bit quantization strategy for \( X_i \) at Sensor \( i \) can be expressed in terms of the conditional probability

\[
P_i(m|x) := p_{M_i|X_i}(m|x) \quad \text{for} \quad m \in [1, 2^k] \quad \text{and} \quad x \in X. \tag{5}
\]

The sensors communicate their \( k \)-bit quantization messages simultaneously and independently to a fusion center whose goal is to produce an estimate of the unknown distribution \( f(x|\theta) \) from the tuple \( M^{(n)} = (M_1, \ldots, M_n) \). The fusion center first finds an estimate \( \hat{\theta} := \hat{\theta}(M^{(n)}) \) of the unknown parameter \( \theta \); and then approximates the unknown source distribution as \( f(x|\hat{\theta}) \). Let \( p \in \mathbb{R}, p > 1 \), be given. Our goal is to design the estimator \( \hat{\theta} \) so as to minimize the worst case power-\( p \) Wasserstein risk

\[
\inf_{\theta} \sup_{\hat{\theta} \in \Theta} \mathbb{E} \left[ W_p^p \left( f(x|\hat{\theta}), f(x|\theta) \right) \right] \tag{6}
\]

where the Wasserstein distance between distributions under distance \( d(\cdot, \cdot) \) is defined as in [1]. As we already mentioned, when the distance \( d(\cdot, \cdot) \) is the \( L_p \)-norm, we also consider the following parameter estimation problem under the \( L_p \)-norm,

\[
\inf_{\theta} \sup_{\hat{\theta} \in \Theta} \mathbb{E} \left[ \left\| \hat{\theta} - \theta \right\|_p^p \right]. \tag{7}
\]

We assume that for all \( i = 1, \ldots, n \) there is a well defined joint probability distribution with density

\[
f_i(x, m|\theta) = f(x|\theta)p_i(m|x) \tag{8}
\]

and that \( p_i(m|x) \) is a regular conditional probability (it denotes the encoding function at the \( i \)-th Sensor – see [5]). For a given \( \theta \in \mathbb{R}^d \) and quantization strategy at the \( i \)-th Sensor, the likelihood that the quantization message \( M_i \) takes a specific value \( m \) is denoted as \( p_i(m|\theta) \). The vector

\[
S_i, \theta(m) = (S_{i,\theta_1}(m), \ldots, S_{i,\theta_d}(m)) = \left( \frac{\partial}{\partial \theta_1} \log p_i(m|\theta), \ldots, \frac{\partial}{\partial \theta_d} \log p_i(m|\theta) \right) \tag{9}
\]

is the score function of this likelihood. For convenience, for \( x \in X' \) we let

\[
S_\theta(x) = (S_{\theta_1}(x), \ldots, S_{\theta_d}(x)) = \left( \frac{\partial}{\partial \theta_1} \log f(x|\theta), \ldots, \frac{\partial}{\partial \theta_d} \log f(x|\theta) \right) \tag{10}
\]

denote the score of the likelihood \( f(x|\theta) \).

We make the following assumptions which we assume to hold throughout unless otherwise stated. The distributions \( f(x|\theta) \) and \( \{p_i(m|\theta)\}_{i=1}^n \) are all assumed to be continuously differentiable at every coordinate of \( \theta \). Also, for all \( i = 1, \ldots, n \) the score function \( S_i, \theta(m) \) as well as its \( p \)-th moment exist. Similarly, for all \( i = 1, \ldots, n \) the generalized Fisher information matrix of order \( p \) for estimating \( \theta_i \) from \( M_i \) and that for estimating it from \( X_i \), both defined as in Definition 1 that follows, are assumed to exist and to be continuous in \( \theta_i \).

**Definition 1.** Let \( p \in \mathbb{R} \) with \( p > 1 \) be given. For a multivariate random variable \( X \) with probability distribution \( f(x|\theta) \) that depends on an unknown vector parameter \( \theta = [\theta_1, \ldots, \theta_d] \in \mathbb{R}^d \), for all \( i = 1, \ldots, n \) the generalized Fisher information matrix of order \( p \) for estimating \( \theta_i \) from \( X_i \) is defined as \([12], [15]\)

\[
I^{(p)}(x_i) = \left( \mathbb{E} \left[ \left| \frac{\partial}{\partial \theta_i} \log f(X_i|\theta) \right|^{p-1} \right] \right)^{\frac{1}{p-1}}. \tag{11}
\]

Also, define

\[
\Omega^{(p)}(\theta) := \sum_{i=1}^d \left( \mathbb{E} \left[ \left| \frac{\partial}{\partial \theta_i} \log f(X_i|\theta) \right|^{p-1} \right] \right)^{\frac{1}{p-1}} \tag{12}
\]

which can be interpreted as the trace of the generalized Fisher information matrix of order \( p > 1 \) for estimating \( \theta \) from \( X \).

It can easily be checked that for \( p = 2 \), the quantity \( \Omega^{(2)}(\theta) \) is the trace of the standard Fisher information matrix, i.e., \( \Omega^{(p)}(\theta) = \text{Tr} \left( I^{(2)}(x_i) \right) \). As it will become clearer from the rest of this paper, throughout we will make extensive usage of the quantity \( \Omega^{(p)}(\theta) \) as defined by (12). For example, for the problem of estimating \( \theta \) from the quantization tuple \( M^{(n)} = (M_1, \ldots, M_n) \) we will use

\[
\Omega^{(p)}(M^{(n)}) := \sum_{i=1}^d \left( \mathbb{E} \left[ \left| \frac{\partial}{\partial \theta_i} \log p(M^{(n)}|\theta) \right|^{p-1} \right] \right)^{\frac{1}{p-1}} \tag{13}
\]

where \( p(M^{(n)}|\theta) = \prod_{i=1}^n p_i(M_i|\theta) \) due to the independence of the samples and encoding functions at the sensors. Likewise, for a single quantization message \( M_j, j = 1, \ldots, n \), we use \( \Omega^{(p)}(M_j) \) which is given by the RHS of (13) in which \( p(M^{(n)}|\theta) \) is replaced by \( p_i(M_j|\theta) \). Also, when we take a Bayesian approach and let \( \mu(\theta) \) be a prior on \( \Theta \), we will use

\[
\Omega^{(p)}(\mu) := \sum_{i=1}^d \left( \mathbb{E} \left[ \left| \frac{\partial}{\partial \theta_i} \log \mu(\theta) \right|^{p-1} \right] \right)^{\frac{1}{p-1}} \tag{14}
\]
III. A VAN TREES TYPE INEQUALITY FOR $L_p$-NORMS

In this section, we take a Bayesian approach. We let the parameter space $\Theta$ be the Cartesian product of closed intervals on the real line, i.e., $\Theta = \prod_{i=1}^d [\theta_{i,\min}, \theta_{i,\max}]$. Let $\pi$ some probability distribution on $\Theta$ with a density measure $\mu$ with respect to the Lebesgue measure (a prior on $\theta$). We make the assumption that $\mu(\theta)$ factorizes as $\mu(\theta) = \prod_{i=1}^d \mu_i(\theta_i)$. Also, suppose that $f(x_i)$ and $\mu(\cdot)$ are both absolutely continuous; and that $\mu$ converges to zero at the boundaries of $\Theta$, i.e., for all $i = 1, \ldots, d$,

$$
\lim_{\theta_i \to \theta_{i,\min}} \mu_i(\theta_i) = \lim_{\theta_i \to \theta_{i,\max}} \mu_i(\theta_i) = 0.
$$

For scalar $X$ and $\theta$ (i.e., $d = 1$), the usual van Trees inequality [10], which is a Bayesian version of the well-known Cramér-Rao inequality established for the Euclidean norm $L_2$, states that

$$
\mathbb{E}[(\hat{\theta}(X) - \theta)^2] \geq \frac{1}{\mathbb{E}_\pi[I_X(\theta)] + I(\mu)}
$$

where $I_X(\theta)$ is the standard Fisher information for estimating $\theta$ from $X$ and $I(\mu)$ designates that from the prior.

The following theorem provides a lower bound on the average error in estimating $\theta = (\theta_1, \ldots, \theta_d)$ from $X$ under the $L_p$-norm, for arbitrary $p > 0$. It can be seen a van Trees type inequality for $L_p$ norms. The result can also be regarded as a Bayesian version of one in [14]. Its proof is essentially based on a judicious application of Hölder inequality and is different from the one of [14].

**Theorem 1.** For $p > 1$, the average estimation error under the norm $L_p$ satisfies the following:

i) If $1 < p < 2$, then we have

$$
\mathbb{E} \left[ \left| \hat{\theta}(X) - \theta \right|^p \right] \geq d^p \left( \mathbb{E} \left[ \left( \Omega_X^{(p)}(\theta) \right)^{\frac{p-1}{p}} \left( \Omega^{(p)}(\mu) \right)^{\frac{1}{p}} \right] \right)^{\frac{p-1}{p}}.
$$

ii) If $p \geq 2$, then we have

$$
\mathbb{E} \left[ \left| \hat{\theta}(X) - \theta \right|^p \right] \geq \frac{d^{1+\frac{p-1}{p}}}{\mathbb{E} \left[ \left( \Omega^{(p)}(\mu) \right) \right]}.\frac{d^{p-1}}{\mathbb{E} \left[ \left( \partial_\theta \Omega_X^{(p)}(\theta) \right) \right]}.
$$

**Proof.** The proof of Theorem 1 is given in Section VI-A.

**Remark 1.** It is easy to see that for $p = 2$ the result of Theorem 1 is the standard van Trees inequality [7] (see also [16]). Also, observe that for values of $p \in \mathbb{R}$ which are such that $1 < p < 2$ the result involves generalized Fisher information of order $p$ for both $X$ and the prior $\mu$, whereas for $p \geq 2$ it involves standard Fisher information (i.e. of order 2). We note that for $p \geq 2$, it is possible to derive a bound that is similar to the RHS of (17), i.e., one that involves generalized Fisher information of order $p$, as below

$$
\mathbb{E} \left[ \left| \hat{\theta}(X) - \theta \right|^p \right] \geq \frac{d^{p-1}}{\mathbb{E} \left[ \left( \Omega^{(p)}(\mu) \right) \right]}.
$$

The proof of the lower bound (19) is given in Section VII-B. However, such proof does not seem to compare easily with the RHS of (18).

In addition, the RHS of (18) turns out to be more tractable analytically for the examples that we will consider in the rest of this paper.

**Corollary 1.** For any vector-valued function $\psi(\theta)$ which is continuously differentiable in each component $\psi_i(\theta)$, the following holds:

i) If $1 < p < 2$, we have

$$
\mathbb{E} \left[ \left| \psi(\hat{\theta}(X)) - \psi(\theta) \right|^p \right] \geq \frac{\sum_i \mathbb{E} \left[ \left| \psi_i(\theta) \right|^p \right]}{\left( \mathbb{E} \left[ \left( \Omega_X^{(p)}(\theta) \right) \right] \right)^{\frac{p-1}{p}}}.
$$

ii) If $p \geq 2$, we have

$$
\mathbb{E} \left[ \left| \psi(\hat{\theta}(X)) - \psi(\theta) \right|^p \right] \geq \frac{d^{1+\frac{p-1}{p}}}{\mathbb{E} \left[ \left( \Omega_X^{(p)}(\theta) \right) \right]}.
$$

**Proof.** The proof of Corollary 1 is given in Section VII-C.

IV. DISTRIBUTED PARAMETER ESTIMATION FROM QUANTIZED SAMPLES

Let us now consider the minimax parameter estimation problem [7] described in Section III. Let $p$ be a prior on $\theta$ that factorizes as in Section III and satisfies (15). Substituting $X$ in Theorem 1 with $M^{(n)} = (M_1, \ldots, M_n)$ we obtain a lower bound on the worst case error under the $L_p$ norm. Such bound, however, does not seem to reflect the right behavior for the error decrease as a function of the number of samples $n$ (for given $p > 1$ and fixed $k \geq 1$ and $d \geq 1$). A better bound, which uses the techniques of the proof of Theorem I and combines them appropriately with Marcinkiewicz-Zygmund inequality [18], is stated in the following theorem.

**Theorem 2.** For $p > 1$, the worst case estimation error under the norm $L_p$ satisfies the following:

i) If $1 < p < 2$, we have

$$
\sup_{\theta \in \mathcal{M}^{(n)}} \mathbb{E} \left[ \left| \hat{\theta}(M^{(n)}) - \theta \right|^p \right] \geq \frac{d^{p-1}}{\mathbb{E} \left[ \left( \Omega^{(p)}(\mu) \right) \right]}.
$$

where, for $j = 1, \ldots, n$, $\Omega^{(p)}_{M_j}(\theta)$ is obtained using (13) and $\Omega^{(p)}(\mu)$ is given by (14).
ii) If $p \geq 2$, we have

$$
\sup_{\theta \in \Theta} E_M(\|\hat{\theta}(M^n) - \theta\|_p^p \mid \Theta) \geq d^{(1 + \frac{2}{p})} \left( \sum_{j=1}^n E_{\Theta} \left[ \text{Tr}(I_{M_j}(\theta)) \right] + \text{Tr}(I(\mu)) \right)^{-\frac{2}{p}}.
$$

Proof: 1) Case $1 \leq p < 2$: Let $q \in \mathbb{R}$ such that $\frac{1}{p} + \frac{1}{q} = 1$, i.e., $q = p/(p-1)$. Consider the following two functions $g(\cdot)$ and $h(\cdot)$ defined, for $x \in X$, $\theta = [\theta_1, \ldots, \theta_d] \in \Theta$ and a specific quantization messages tuple $m^{(n)} = (m_1, \ldots, m_n) \in [1, 2^k]^n$ as

$$
g(m^{(n)}, \theta) = \sum_{i=1}^d \frac{\partial}{\partial \theta_i} \left[ \log \left( p(m^{(n)} \mid \theta) \mu(\theta) \right) \right] \tag{20a}
$$

$$
h(m^{(n)}, \theta) = \hat{\theta}(m^{(n)}) - \theta \tag{20b}
$$

where in (20a) the quantization messages joint probability is $p(m^{(n)} \mid \theta) = \prod_{j=1}^n p_j(m_j \mid \theta)$. For convenience, for $i = 1, \ldots, d$ we will denote the $i$th component of $h(m^{(n)}, \theta)$ as $h_i(m^{(n)}, \theta)$, i.e.,

$$
h_i(m^{(n)}, \theta) = \hat{\theta}_i(m^{(n)}) - \theta_i = \left( \hat{\theta}(m^{(n)}, \theta) \right)_i. \tag{21}
$$

Using the fact that the prior measure $\mu$ converges to zero at the boundaries of $\Theta$, it is easy to see that

$$
\sum_{m^{(n)}, \theta} \int h_i(m^{(n)}, \theta) \frac{\partial}{\partial \theta_i} \left[ p(m^{(n)} \mid \theta) \mu(\theta) \right] d\theta_i = 1. \tag{22}
$$

By partial integration and (22), we get for $i = 1, \ldots, d$, that

$$
E_M(\|h_i(M^{(n)}, \Theta)g(M^{(n)}, \Theta)\|_1) = d. \tag{23}
$$

Thus, for all $i = 1, \ldots, d$, we have

$$
d \leq E_M(\|h_i(M^{(n)}, \Theta)g(M^{(n)}, \Theta)\|_q). \tag{24}
$$

Applying Hölder’s inequality for expectations yields

$$
E_M(\|h_i(M^{(n)}, \Theta)g(M^{(n)}, \Theta)\|_q) \leq \left( E \left[ \|h_i(M^{(n)}, \Theta)\|_p \right] \right)^\frac{1}{q} \left( E \left[ \|g(M^{(n)}, \Theta)\|_q \right] \right)^\frac{1}{q}. \tag{25}
$$

The first element of the right-hand side produces the desired risk as

$$
\sup_{\theta \in \Theta} E_M(\Theta) \left[ \left\| \hat{\theta}(M^{(n)}) - \theta \right\|_p^p \mid \Theta \right] \geq \sum_{i=1}^d E_M(\Theta) \left[ h_i(M^{(n)}, \Theta) \right]^p \tag{26}
$$

where the inequality follows by substituting using (21) and using that fact that the supremum of a function is larger than its expectation.

We now upper bound the second expectation term of the RHS of (26). For convenience, let for $j = 1, \ldots, 2^k$

$$
l(m_j, \theta) = \sum_{i=1}^d \frac{\partial}{\partial \theta_i} \left[ \log p(m_j \mid \theta) \right]. \tag{28}
$$

It is easy to see that for all $\theta$, we have

$$
E_{M_j \mid \Theta} \left[ l(M_j, \Theta) \mid \Theta = \theta \right] = 0. \tag{29}
$$

Then, we have

$$
\left( E_M(\Theta) \left[ ||g(M^{(n)}, \Theta)||_q \right]^q \right)^\frac{1}{q} \leq \left( E_M(\Theta) \left[ \left\| \hat{\theta}(M^{(n)}) - \theta \right\|_p^p \right] \right)^\frac{1}{p} + d^{\frac{p-1}{p}} (\Omega_p(\mu))^\frac{1}{p}, \tag{30}
$$

where the inequality holds by a double application of Minkowski’s inequality: first for expectations using that for all $Z$ and $T$ we have $E[Z + T]^q \leq E[Z]^q + E[T]^q$; and then that $E[\sum_{i=1}^d |Z_i|^q] \leq \sum_{i=1}^d (E[|Z_i|^q])$ for all $q > 1$ and $d \geq 1$, the application of Marcinkiewicz-Zygmund inequality yields

$$
E_M(\Theta) \left[ \left( \sum_{j=1}^n l(M_j, \Theta) \right)^q \mid \Theta \right] \leq B_q E_M(\Theta) \left[ \left( \sum_{j=1}^n l^2(M_j, \Theta) \right)^{\frac{q}{2}} \mid \Theta \right], \tag{31}
$$

where $B_q = (q-1)^{\frac{q}{2}} > 0$. Continuing from (31), we get

$$
\left( E_M(\Theta) \left[ \left\| l(M_j, \Theta) \right\|_1 \right] \right)^{\frac{q}{p}} \leq \left( \frac{1}{(p-1)^2} \sum_{j=1}^n \left( E_M(\Theta) \left[ \left\| l(M_j, \Theta) \right\|_1 \right] \right)^q \right)^{\frac{p-1}{p}} \tag{32a}
$$

$$
\leq \left( \frac{d^{\frac{p}{p-1}}}{(p-1)^2} \sum_{j=1}^n \left( E_M(\Theta) \left[ \left( \frac{d^{\frac{p-1}{p}} \sum_{u_{i_1} \in \mathcal{U}_j} \left( \log p(M_j \mid \theta) \right)^\frac{1}{p} \right)^\frac{2(p-1)}{q} \right) \right] \right)^\frac{\frac{p-1}{p}}{2}, \tag{32b}
$$

where (a) follows by using the quantization messages are independent, substituting $q = p/(p-1)$ and applying Minkowski’s inequality $E[\sum_{i=1}^n X_i]^q \leq \sum_{i=1}^n (E[X_i]^q)$ since $q = p > (p-1) > 2$; (b) follows by substituting using (28) and using that $\sum_{i=1}^d u_i \leq d^{\frac{p-1}{p}} \sum_{i=1}^d u_i^q, u_i > 0$; and (c) holds by (13). Finally, combining (22), (30), (24) and (27) and substituting in (26) yields the desired result.

2) Case $p \geq 2$: In this case, a direct proof can be found in a way that is essentially similar to the above (see Section VI-D1 for the details). An indirect proof follows by first observing that

$$
\left( E_M(\Theta) \left[ \left\| \hat{\theta}(M^{(n)}) - \theta \right\|_p^p \right] \right)^\frac{1}{p} \geq d^{1-\frac{p}{2}} \left( E_M(\Theta) \left[ \left\| \hat{\theta}(M^{(n)}) - \theta \right\|_2^2 \right] \right)^\frac{1}{2}, \tag{33}
$$

which holds due to the norms inequality $\|u\|_2 \leq d^{\frac{p-1}{p}} \|u\|_p$ for all vector $u \in \mathbb{R}^d$; and then combining with the result of (2) for the squared $L_2$ loss.
Corollary 2. For the special case $p = 2$, the result of Corollary 2 recovers that of [2, Corollary 5].

V. Estimation under the Wasserstein Loss

We now turn to the minimax risk given by (6) in Section II. Theorem 3 of Section [V] as well as its proof, are instrumental in obtaining similar bounds for the Wasserstein loss (9) when the underlying distance $d(\cdot, \cdot)$ is based on the $L_p$-norm. For the Gaussian location model (see Corollary 3 below) this yields a lower bound on the worst-case Wasserstein loss under the $L_p$-norm which decreases at least as $n^{-\frac{p}{p+1}}$.

Theorem 4. For any estimator $\hat{\theta} = \hat{\theta}(M(n))$, the following holds.

i) If $1 < p < 2$, we have

$$\sup_{\theta \in \Theta} \mathbb{E}_{M(n) | \Theta} \left[ W_p^p(f(x | \hat{\theta}(M(n))), f(x | \theta)) | \Theta \right] \geq \frac{1}{d^p} \int_0^1 \int_{B_p,d} \left[ \partial_{\theta} \mathbb{E}_{\mathcal{Y} - f(y | \theta) | Y_j]} \right] \left. \right|_{\partial Y = Y_j} \left[ \left( \Omega_p(Y_j | \Theta) \right) \right]^{\frac{1}{2}} \left( Y_j \right) \right]^{\frac{2(p-1)}{p}} + d \left( \frac{d-1}{2p} \right) \left( \Omega_p(Y_j | \Theta) \right)^{\frac{1}{p}} \right]^{\frac{1}{p}} \left( \Omega_p(Y_j | \Theta) \right)^{\frac{1}{p}} \right]^{\frac{1}{p}} \right.$$
A. Proof of Theorem 1

By partial integration and (42), we get for $\Theta$ optimal rates of the order of the upper bound in [21] for the empirical estimator smoothed by that of the order $n^{-1}$.

Proof. The proof of Corollary 3 is given in Section VI-H.

A $K$-subgaussian distribution, estimated with an empirical distribution smoothed by a Gaussian kernel, enjoys upper bounds on the error in the 1-Wasserstein distance, $W_1$, of the order $n^{-\frac{1}{2}}$ and in the squared 2-Wasserstein distance, $W_2^2$, of the order $n^{-1}$ [21]. The bounds show remarkable performance improvement of this convolution over the unsmoothed empirical estimator from $n^{-\frac{1}{2}}$ to that of the order $n^{-\frac{1}{2}}$ for $W_1$ and $n^{-1}$ for the $W_2^2$. If $p = 2$, we obtain a lower bound on $W_2^2$ of the order $n^{-1}$, which matches that of the upper bound in [21] for the empirical estimator smoothed by a Gaussian kernel of a $K$-subgaussian distribution. Our technique may be useful in [21], to produce a matching lower bound, to yield optimal rates of the order $n^{-1}$.

VI. PROOFS

A. Proof of Theorem 7

Let $g \in \mathbb{R}$ such that $\frac{1}{p} + \frac{1}{q} = 1$, i.e., $q = p/(p-1)$. Also, consider the following two functions $g(\cdot)$ and $h(\cdot)$ defined, for $x \in \mathcal{X}$ and $\theta = [\theta_1, \ldots, \theta_d] \in \Theta$, as

\[
g(x, \theta) = \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \log (f(x|\theta)\mu(\theta))] \quad (40a)
\]

\[
h(x, \theta) = \hat{h}_i(x) - \theta_i = (h(x, \theta))_i \quad (41)
\]

Using the fact that the prior measure $\mu$ converges to zero at the endpoints of $\Theta$, it is easy to see that

\[
\int_{\mathcal{X}} h_i(x, \theta) \frac{\partial}{\partial \theta_i} [f(x|\theta)\mu(\theta)] \, dx = 1. \tag{42}
\]

By partial integration and (42), we get for $i = 1, \ldots, d$, that

\[
E_{(X, \Theta)} [h_i(X, \Theta)g(X, \Theta)] = d. \tag{43}
\]

Thus, for all $i = 1, \ldots, d$, we have

\[
E_{(X, \Theta)} [h_i(X, \Theta)g(X, \Theta)] \geq d. \tag{44}
\]

For convenience, let

\[
l(x, \theta) = \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \log f(x|\theta). \tag{45}
\]

It is easy to see that for all $\theta$, we have

\[
E_{X|\Theta} [l(X, \Theta)|\Theta = \theta] = 0 \tag{46}
\]

where follows by the regularity condition

\[
E_{X|\Theta} \left[ \frac{\partial}{\partial \theta_i} \log f(x|\theta) \right] = 0 \text{ for all } \theta \in \Theta. \]
Thus,
\[
E_{\Theta} \left( \left[ E_{X|\Theta} \left[ |g(X, \Theta)|^2 \right] \right]^\frac{2}{q} \right) \leq \left( E \left[ |g(X, \Theta)|^2 \right] \right)^\frac{2}{q} \tag{54}
\]
\[
\leq \left( E_{X, \Theta} \left[ |g(X, \Theta)|^2 \right] \right)^\frac{2}{q} + E_{\Theta} \left( \sum_{i=1}^d \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2_{\frac{2}{q}} \tag{55}
\]
where (a) follows using Jensen’s inequality for the concave function \( u \to u^{\frac{q}{2}} \) for \( q = p/p-1 \leq 2; \) and (b) follows by substituting using (53).

The first expectation term on the RHS of (55) is upper bounded as
\[
E_{X, \Theta} \left[ |g(X, \Theta)|^2 \right] \leq \left( E \left[ |g(X, \Theta)|^2 \right] \right)^\frac{2}{q} \tag{56}
\]
where (a) follows by substituting using (44) and (b) holds since for non-negative \((u_i)_{i=1}^d\) we have \( \left( \sum_{i=1}^d u_i \right)^{\frac{q}{2}} \leq d \sum_{i=1}^d u_i^q \).

Hence, we get
\[
\left( E_{\Theta} \left[ \left( E_{X|\Theta} \left[ |g(X, \Theta)|^2 \right] \right)^\frac{2}{q} \right] \right)^\frac{2}{q} \leq \left( E_{X, \Theta} \left[ |g(X, \Theta)|^2 \right] \right)^\frac{2}{q} + E_{\Theta} \left( \sum_{i=1}^d \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2_{\frac{2}{q}} \tag{57}
\]
where (a) follows by using (55) and noticing that \( p/q = p-1 \geq 1 \) and (b) holds using (56).

Summarizing, combining (55) and (57) we get
\[
E_{X, \Theta} \left[ \left| \hat{\theta}(X) - \theta \right|^p \right] \geq \frac{d^{1+\frac{q}{2}}}{\left( E_{\Theta} \left[ \text{Tr}(I_X(\Theta)) \right] + d \text{Tr}(I(\mu)) \right)^\frac{q}{2}} \tag{58}
\]

2) Case \( 1 < p < 2): \) First, recall that
\[
E_{X, \Theta} \left[ \left| \hat{\theta}(X) - \theta \right|^p \right] = \sum_{i=1}^d E_{X, \Theta} \left[ |h_i(X, \Theta)|^p \right] \tag{59}
\]

Also, for all \( i = 1, \ldots, d, \) an easy application of Hölder’s inequality for expectations yields
\[
E_{X, \Theta} \left[ |h_i(X, \Theta) g(X, \Theta)| \right] \leq \left( E_{X, \Theta} \left[ |h_i(X, \Theta)|^p \right] \right)^{\frac{1}{p}} \left( E_{X, \Theta} \left[ |g(X, \Theta)|^q \right] \right)^{\frac{1}{q}} \tag{60}
\]

Thus, using (59) and (60), we get
\[
E_{X, \Theta} \left[ \left| \hat{\theta}(X) - \theta \right|^p \right] \geq \frac{d^{p+1}}{\left( E_{X, \Theta} \left[ |g(X, \Theta)|^q \right] \right)^{\frac{1}{q}}} \tag{61}
\]

The rest of the proof in this case is devoted to upper-bounding the denominator of the RHS of (61).

Recalling (47), we have
\[
\left( E_{X, \Theta} \left[ \left| g(X, \Theta) \right|^q \right] \right)^{\frac{1}{q}} \leq \left( E_{X, \Theta} \left[ \left| g(X, \Theta) \right|^p \right] \right)^{\frac{1}{p}} \tag{62}
\]
\[

\left( E_{X, \Theta} \left[ \left| g(X, \Theta) \right|^q \right] \right)^{\frac{1}{q}} \leq \left( E_{X, \Theta} \left[ \left| g(X, \Theta) \right|^p \right] \right)^{\frac{1}{p}} + \left( \sum_{i=1}^d \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2_{\frac{2}{q}} \tag{63}
\]

where: (a) follows by applying the Minkowski’s inequality for expectations \((E||Z + T\\|q|^q)^{\frac{1}{q}} \leq (E||Z\\|q|^q)^{\frac{1}{q}} + (E||T\\|q|^q)^{\frac{1}{q}}\) for r.v.s \( Z \) and \( T \); (b) follows by application of the Minkowski’s inequality for expectations \((E\sum_{i=1}^d \{E|Z_i\\|q|^q\})^{\frac{1}{q}} \leq \sum_{i=1}^d \{E|Z_i\\|q|^q\}^{\frac{1}{q}}\) for r.v.s \( Z_1, \ldots, Z_d \); (c) follows by substituting using \( q = p/p-1 \) and (d) holds by first using the inequality \( \sum_{i=1}^d u_i^q \leq d^{\frac{p-1}{p}} \left( \sum_{i=1}^d u_i^p \right)^{\frac{q}{p}} \) for non-negative \((u_1, \ldots, u_d)\) and \( p > 1 \) and then substituting using (44).

Continuing from (63), the first term of its RHS can be upper bounded as
\[
\left( E_{X, \Theta} \left[ \left| g(X, \Theta) \right|^q \right] \right)^{\frac{1}{q}} \leq d^{p+1} \left( E_{X, \Theta} \left[ \left| g(X, \Theta) \right|^p \right] \right)^{\frac{1}{p}} \tag{64}
\]

where: (a) follows by substituting using (44); (b) holds by using the inequality \( \left( \sum_{i=1}^d \sum_{i=1}^d u_i^q \right)^{\frac{1}{p}} \leq d^{\frac{p-1}{p}} \sum_{i=1}^d u_i^q \) which holds for non-negative \((u_1, \ldots, u_d)\) and \( q > 1 \), (c) follows by substituting using \( q = \frac{p}{p-1} \); and (d) holds by defining, for \( i = 1, \ldots, d \) and \( \Theta \in \Theta, \)
\[
\nu_i(\Theta) = \left( E_{X|\Theta} \left[ \frac{\partial}{\partial \theta_i} \log f(X|\Theta) \right] \right)^{\frac{1}{p}} | \Theta = \theta \right)^{p-1} \tag{65}
\]
(c) holds by using the inequality \( \sum_{i=1}^{d} u_i^{\frac{1}{p'}} \leq \left( \sum_{i=1}^{d} u_i \right)^{\frac{1}{p'}} \) for non-negative \((u_1, \ldots, u_d)\) and \(p < 2\); and (f) holds by substituting using (12).

Hence, combining (64) and (62), we get

\[
(\mathbb{E}(X, \Theta) \left[ |g(X, \Theta)|^q \right])^{\frac{1}{q}} \leq d^p \left( \mathbb{E}(\Theta) \left[ \left( \Omega_X^{(p)}(\theta) \right)^{\frac{1}{p}} \right] \right)^{\frac{p-1}{p}} + \left( \mathbb{E}(\Theta) \left[ \left( \Omega^A^{(p)}(\mu) \right)^{\frac{1}{p}} \right] \right)^{\frac{p-1}{p}}. \tag{66}
\]

Finally, substituting in (61) using (66) yields the desired result,

\[
\mathbb{E}(X, \Theta) \left[ ||\hat{\theta}(X) - \theta||_p \right] \geq d^p \left( d^{\frac{p}{p'-2}} \left( \Omega^A^{(p)}(\mu) \right)^{\frac{p-1}{p}} \right) + \left( \mathbb{E}(\Theta) \left[ \left( \Omega_X^{(p)}(\theta) \right)^{\frac{1}{p}} \right] \right)^{\frac{p-1}{p}}. \tag{67}
\]

**B. Proof of Inequality (79)**

Let \( q \in \mathbb{R} \) such that \( \frac{1}{p} + \frac{1}{q} = 1 \), i.e., \( q = p/(p-1) \). Also, consider the following two functions \( g(\cdot) \) and \( h(\cdot) \) defined, for \( x \in \mathcal{X} \) and \( \theta = [\theta_1, \ldots, \theta_d] \in \Theta \), as

\[
g(x, \theta) = \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \log(f(x|\theta)) \mu(\theta) ] \tag{68a}
\]

\[
h(x, \theta) = \psi(\hat{\theta}(x)) - \psi(\theta). \tag{68b}
\]

For convenience, for \( i = 1, \ldots, d \) we will denote the \( i \)-th component of \( h_i(x, \theta) \) as \( h_i(x, \theta) \), i.e.,

\[
h_i(x, \theta) = \psi(\hat{\theta}(x)) - \psi(\theta_i) = (h(x, \theta))_i \tag{69}
\]

Using the definition of the \( p \)-norm, the average estimation error can be lower bounded as

\[
\mathbb{E}(X, \Theta) \left[ \left| \left| \psi(\hat{\theta}(X)) - \psi(\theta) \right| \right|_p \right] = \sum_{i=1}^{d} \mathbb{E}(X, \Theta) \left[ |h_i(X, \Theta)|^p \right]. \tag{70}
\]

The RHS of (70) can be lower bounded as follows. First, note that applying Hölder’s inequality for expectations yields

\[
\mathbb{E}(X, \Theta) \left[ |h_i(X, \Theta)| g(X, \Theta) \right] \leq \left( \mathbb{E}(X, \Theta) \left[ |h_i(X, \Theta)|^p \right] \right)^{\frac{1}{p}} \times \left( \mathbb{E}(X, \Theta) \left[ |g(X, \Theta)|^q \right] \right)^{\frac{1}{q}}. \tag{71}
\]

Using the fact that the prior measure \( \mu \) converges to zero at the endpoints of \( \Theta \) and partial integration, it is easy to see that

\[
\int_{\Theta} h_i(x, \theta) \frac{\partial}{\partial \theta_i} \left[ f(x|\theta) \mu_i(\theta) \right] \, d\theta_i = \int_{\Theta} f(x|\theta) \mu_i(\theta) \left( g_i(x) \right)_{\text{max-min}} \, d\theta_i \tag{80}
\]

\[
\int_{\Theta} \frac{\partial}{\partial \theta_i} \left[ h_i(x, \theta) \right] \, f(x|\theta) \mu_i(\theta) \, d\theta_i = - \int_{\Theta} \frac{\partial}{\partial \theta_i} \left[ h_i(x, \theta) \right] \, f(x|\theta) \mu_i(\theta) \, d\theta_i. \tag{72}
\]

Integration in (72), we get for \( i = 1, \ldots, d \), that

\[
\int_{\mathcal{X}} \int_{\Theta} h_i(x, \theta) \frac{\partial}{\partial \theta_i} \left[ f(x|\theta) \mu_i(\theta) \right] \, d\theta_i \, dx = - \mathbb{E}(X, \Theta_i) \left[ \frac{\partial}{\partial \theta_i} \left[ h_i(x, \theta) \right] \right]. \tag{73}
\]

Thus, with some algebraic manipulations,

\[
\mathbb{E}(X, \Theta) \left[ |h_i(X, \Theta)| g(X, \Theta) \right] = \sum_{i=1}^{d} \mathbb{E}(\Theta_i) \left[ \left| \frac{\partial}{\partial \theta_i} h_i(X, \Theta) \right| \right]. \tag{74}
\]

Combining (70), (71) and (75), we get

\[
\mathbb{E}(X, \Theta) \left[ \left| \psi(\hat{\theta}(X)) - \psi(\theta) \right| \right] \geq d^p \left( d^{\frac{p}{p'-2}} \left( \Omega^A^{(p)}(\mu) \right)^{\frac{p-1}{p}} \right) + \left( \mathbb{E}(\Theta) \left[ \left( \Omega_X^{(p)}(\theta) \right)^{\frac{1}{p}} \right] \right)^{\frac{p-1}{p}}. \tag{76}
\]

We now upper bound the second expectation of the RHS term of (76), as follows. For convenience, let

\[
l(x, \theta) = \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \log f(x|\theta). \tag{77}
\]

It is easy to see that for all \( \theta \), we have

\[
\mathbb{E}(X|\Theta) \left[ l(X, \Theta) \right] = 0 \tag{78}
\]

which follows by the regularity condition \( \mathbb{E}(X|\Theta) \left[ \frac{\partial}{\partial \theta_i} \log f(x|\theta) \right] = 0 \) for all \( \theta \in \Theta \). Also,

\[
g(x, \theta) = l(x, \theta) + \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \log \mu(\theta). \tag{79}
\]

Note that \( l(x, \theta) \) is the sum of the elements of the score function associated with \( X \).

From (76), we have

\[
\mathbb{E}(X, \Theta) \left[ |g(X, \Theta)|^q \right] \leq \left( \mathbb{E}(X, \Theta) \left[ |h_i(X, \Theta)|^p \right] \right)^{\frac{1}{p}} \times \left( \mathbb{E}(X, \Theta) \left[ |g(X, \Theta)|^q \right] \right)^{\frac{1}{q}}. \tag{80}
\]

Therefore,

\[
\mathbb{E}(X, \Theta) \left[ |l(X, \Theta)|^q \right] \geq \left( \mathbb{E}(X, \Theta) \left[ |h_i(X, \Theta)|^p \right] \right)^{\frac{1}{p}} + \left( \mathbb{E}(X, \Theta) \left[ \left| \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right|^q \right] \right)^{\frac{1}{q}}. \tag{81}
\]

where: (a) follows by application of the Minkowski’s inequality for expectations \((E[|Z + T|^p])^{\frac{1}{p}} \leq (E[|Z|^p])^{\frac{1}{p}} + (E[|T|^p])^{\frac{1}{p}} \) for r.v.s \( Z \).
and \( T \); (b) follows by application of the Minkowski’s inequality for expectations \( \left( \mathbb{E}\left[ \sum_{i=1}^{d} Z_i \right] \right)^{\frac{1}{p}} \leq \sum_{i=1}^{d} \left( \mathbb{E}\left[ Z_i \right] \right)^{\frac{1}{p}} \) for r.v.s \((Z_1, \ldots, Z_d)\); (c) holds by substituting \( q = p/p - 1 \) and (d) holds by first using the inequality \( \sum_{i=1}^{d} u_i \geq d^{p-1} \left( \sum_{i=1}^{d} u_i \right)^{\frac{1}{p}} \) for non-negative \((u_1, \ldots, u_d)\) and \( p > 1 \) and then substituting using (12).

Continuing from (81), the first term of its RHS can be upper bounded as

\[
\left( \mathbb{E}_{(X, \Theta)} \left[ [l(X, \Theta)]^{q} \right] \right)^{\frac{1}{q}} \leq \left( \mathbb{E}_{(X, \Theta)} \left[ \left( \frac{d^{p-1}}{\sum_{i=1}^{d} u_i} \left[ \mathbb{E}(f(X|\Theta)) \right]^{\frac{p}{p-1}} | \Theta = \theta \right) \right] \right)^{\frac{1}{p-1}} \leq \left( \frac{d^{\frac{p-1}{p}}}{\sum_{i=1}^{d} u_i} \left[ \mathbb{E}(f(X|\Theta)) \right]^{\frac{p}{p-1}} | \Theta = \theta \right)^{\frac{1}{p-1}},
\]

(82)

where: (a) follows by substituting using (72); (b) holds by using the inequality \( \left( \frac{1}{\sum_{i=1}^{d} u_i} \right)^{q} \leq d^{q-1} \sum_{i=1}^{d} u_i^{q} \) which holds for non-negative \((u_1, \ldots, u_d)\) and \( q > 1 \), (c) follows by substituting using \( q = \frac{p}{p-1} \); and (d) holds by defining, for \( i = 1, \ldots, d \) and \( \Theta, \theta \in \Omega \),

\[
v_i(\theta) = \left( \mathbb{E}_{(X, \Theta)} \left[ \left[ \frac{\partial}{\partial \theta_i} \log f(X|\Theta) \right]^{\frac{p}{p-1}} | \Theta = \theta \right) \right)^{\frac{1}{p-1}} ; \quad (83)
\]

\( (e) \) holds by using the inequality \( \sum_{i=1}^{d} u_i \leq d^{\frac{p-1}{p}} \left( \sum_{i=1}^{d} u_i \right)^{\frac{1}{p}} \) for non-negative \((u_1, \ldots, u_d)\) and \( p \geq 2 \); and \((f)\) holds by substituting using (12).

Substituting (82) in (81), we obtain

\[
\left( \mathbb{E}_{(X, \Theta)} \left[ [g(X, \Theta)]^{q} \right] \right)^{\frac{1}{q}} \leq d^{\frac{p-1}{p}} \left( \mathbb{E}_{(X, \Theta)} \left[ \left( \frac{\partial}{\partial \theta_i} \log f(X|\Theta) \right)^{\frac{p}{p-1}} | \Theta = \theta \right) \right)^{\frac{1}{p-1}} + d^{\frac{p-1}{p}} \left( \Omega^{(p)}(\mu) \right)^{\frac{1}{p}}.
\]

(84)

Substituting (83) in (76) produces the lower bound

\[
\mathbb{E}_{(X, \Theta)} \left[ [\psi(\Theta) - \psi(\hat{\Theta})]^{p} \right] \geq \mathbb{E}_{\Theta} \left[ \left( \frac{d^{p}}{\sum_{i=1}^{d} u_i} \left[ \mathbb{E}(f(X|\Theta)) \right]^{\frac{p}{p-1}} | \Theta = \theta \right) \right] \left( \Omega^{(p)}(\mu) \right)^{\frac{1}{p}} \leq d^{\frac{p-1}{p}} \left( \Omega^{(p)}(\mu) \right)^{\frac{1}{p}} + \left( \mathbb{E}_{\Theta} \left[ \left( \frac{\partial}{\partial \theta_i} \log f(X|\Theta) \right)^{\frac{p}{p-1}} | \Theta = \theta \right) \right)^{\frac{1}{p-1}}.
\]

(85)

Let \( \psi(\Theta) = \theta \), then we obtain

\[
\mathbb{E}_{(X, \Theta)} \left[ [\theta(X) - \theta]^{p} \right] \geq d^{\frac{p-1}{p}} \left( \Omega^{(p)}(\mu) \right)^{\frac{1}{p}} + \left( \mathbb{E}_{\Theta} \left[ \left( \frac{\partial}{\partial \theta_i} \log f(X|\Theta) \right)^{\frac{p}{p-1}} | \Theta = \theta \right) \right)^{\frac{1}{p-1}}.
\]

C. Proof of Corollary 7

1) Case \( p \geq 2 \): Let \( q \in \mathbb{R} \) such that \( \frac{1}{p} + \frac{1}{q} = 1 \), i.e., \( q = p/(p - 1) \). Also, consider the following two functions \( g(\cdot) \) and \( h(\cdot) \) defined, for \( x \in \mathcal{X} \) and \( \theta = [\theta_1, \ldots, \theta_d] \in \Theta \), as

\[
g(x, \theta) = \frac{d}{\sum_{i=1}^{d} \left[ \log(f(x|\theta)\mu(\theta)) \right]} \quad \forall x \in \mathcal{X} \quad \forall \theta \in \Theta \quad \forall \mu(\theta) \in \mathcal{M}(\Theta)
\]

(85a)

\[
h(x, \theta) = \psi(\hat{\theta}(x)) - \psi(\theta).
\]

(85b)

For convenience, for \( i = 1, \ldots, d \) we will denote the \( i \)th component of \( h_i(x, \theta) \) as \( h_i(x, \theta) \), i.e.,

\[
h_i(x, \theta) = \psi(\hat{\theta}(x)) - \psi(\theta). \quad (86)
\]

The average estimation error can be lower bounded as

\[
\mathbb{E}_{(X, \Theta)} \left[ [\psi(\hat{\Theta}(x)) - \psi(\theta)]^{p} \right] \geq \left( \frac{d}{\sum_{i=1}^{d} \left[ \log(f(x|\theta)\mu(\theta)) \right]} \right)^{\frac{1}{p}} \left( \Omega^{(p)}(\mu) \right)^{\frac{1}{p}} + \left( \mathbb{E}_{\Theta} \left[ \left( \frac{\partial}{\partial \theta_i} \log f(X|\Theta) \right)^{\frac{p}{p-1}} | \Theta = \theta \right) \right)^{\frac{1}{p-1}}.
\]

(87)

where (a) follows from the definition of the \( p \)-norm and (b) by replacing \( \mathbb{E}_{(X, \Theta)} \) with \( \mathbb{E}_{(X, \Theta)} \) and \( \mathbb{E}_{X|\Theta} \) and Jensen’s inequality for expectations for convex functions \( x \mapsto x^{\frac{1}{p}} \), for \( 2 < p \).

The RHS of (88) can be lower bounded as first, note that we have

\[
\mathbb{E}_{(X, \Theta)} \left[ [h_i(X, \Theta)]^{p}(\mathbb{E}(g(X|\Theta)) \Theta = \theta) \right] \geq \mathbb{E}_{(X, \Theta)} \left[ [h_i(X, \Theta)]^{p}(\mathbb{E}(g(X|\Theta)) \Theta = \theta) \right]
\]

(88)

where (a) follows by application of Hölder’s inequality for every \( \Theta \in \Theta \) to the conditional expectation \( \mathbb{E}_{X|\Theta} [\cdot | \Theta] \); and (b) follows by application of Hölder’s inequality to the expectation \( \mathbb{E}_{\Theta} [\cdot] \) since \( p > 1, q > 1 \) and are such that \( \frac{1}{p} + \frac{1}{q} = 1 \).
Using the fact that the prior measure $\mu$ converges to zero at the endpoints of $\Theta$, it is easy to see that

$$
\int_{\Theta_i} h_i(x, \theta) \frac{\partial}{\partial \theta_i} \left[ f(x, \theta) \mu_i(\theta_i) \right] \, d\theta_i
= h_i(x, \theta) f(x, \theta) \mu_i(\theta_i) \frac{\partial}{\partial \theta_i} \sum_{i=1}^{d} | \mu_i(\theta_i) | \, d\theta_i
= - \int_{\Theta_i} \frac{\partial}{\partial \theta_i} [ h_i(x, \theta) ] \, f(x, \theta) \mu_i(\theta_i) \, d\theta_i, \quad (90)
$$

By partial integration and (90), we get for $i = 1, \ldots, d$, that

$$
\int_{\Theta} \int_{\Theta_i} h_i(x, \theta) \frac{\partial}{\partial \theta_i} \left[ f(x, \theta) \mu_i(\theta_i) \right] \, d\theta_i \, dx
= - \mathbb{E}_{(X, \Theta)} \left[ \frac{\partial}{\partial \theta_i} [ h_i(X, \Theta) ] \right]. \quad (91)
$$

Thus, with some algebraic manipulations,

$$
\mathbb{E}_{(X, \Theta)} [ h_i(X, \Theta) g(X, \Theta) ]
= \sum_{i=1}^{d} \mathbb{E}_{\Theta_i} \left[ \mathbb{E}_{X|\Theta} \left[ \frac{\partial}{\partial \theta_i} [ h_i(X, \Theta) ] \right] \right]
= - \sum_{i=1}^{d} \mathbb{E}_{(X, \Theta)} \left[ \frac{\partial}{\partial \theta_i} [ h_i(X, \Theta) ] \right]
= \sum_{i=1}^{d} \mathbb{E}_{(X, \Theta)} \left[ \frac{\partial}{\partial \theta_i} \psi_i(\theta) \right], \quad (92)
$$

and $|E[X]| \leq E[|X|]$ lower bounds the left-hand side of (89)

$$
\left| \mathbb{E}_{(X, \Theta)} \left[ \frac{\partial}{\partial \theta_i} \psi_i(\theta) \right] \right| \leq \mathbb{E}_{(X, \Theta)} [ |h_i(X, \Theta) g(X, \Theta)| ]. \quad (93)
$$

We now upper bound the second expectation of the RHS term of (89), as follows. For convenience, let

$$
l(x, \theta) = \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \left[ \log f(x|\theta) \right]. \quad (94)
$$

It is easy to see that for all $\theta$, we have

$$
\mathbb{E}_{X|\Theta} \left[ l(X, \Theta) \Theta = \theta \right] = 0 \quad (95)
$$

which follows by the regularity condition

$$
\mathbb{E}_{X|\Theta} \left[ \frac{\partial}{\partial \theta_i} \log f(x|\theta) \right] = 0 \text{ for all } \theta \in \Theta. \text{ Also,}
$$

$$
g(x, \theta) = l(x, \theta) + \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \left[ \log \mu(\theta) \right]. \quad (96)
$$

Now, since

$$
\mathbb{E}_{X|\Theta} \left[ l(X, \Theta) \left( \frac{\partial}{\partial \theta_i} \log \mu_i(\theta_i) \right) \right] \Theta = \theta = 0, \quad (97)
$$

we get

$$
\mathbb{E}_{X|\Theta} \left[ g(X, \Theta) \right] \Theta = \theta
= \mathbb{E}_{X|\Theta} \left[ \frac{\partial}{\partial \theta_i} \psi_i(\theta) \right] + \left( \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \left[ \log \mu(\theta) \right] \right)^{2}. \quad (98)
$$

Thus,

$$
\mathbb{E}_{\Theta} \left[ \left( \mathbb{E}_{X|\Theta} \left[ |g(X, \Theta)|^2 \mid \Theta = \theta \right] \right)^{\frac{1}{2}} \right]
\leq \left( \mathbb{E}_{\Theta} \left[ |g(X, \Theta)|^2 \right] \right)^{\frac{1}{2}}, \quad (a)
$$

$$
\leq \left( \mathbb{E}_{(X, \Theta)} \left[ g^2(X, \Theta) \right] + \mathbb{E}_{\Theta} \left[ \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \left[ \log \mu(\theta) \right] \right)^{2} \right)^{\frac{1}{2}}, \quad (b) \quad (99)
$$

where (a) follows using Jensen’s inequality for the concave function $u \mapsto u^{2/2}$ for $q = p/p-1 \leq 2$; and (b) follows by substituting using (98).

The first expectation term on the RHS of (100) is upper bounded as

$$
\mathbb{E}_{(X, \Theta)} \left[ g^2(X, \Theta) \right] \leq \frac{d \mathbb{E}_{\Theta} [ \text{Tr}(I(X(\Theta))) ]}{d} \quad (101)
$$

where (a) follows by substituting using (99), and (b) holds since for non-negative $\{u_i\}_{i=1}^{d}$ we have $\left( \sum_{i=1}^{d} u_i \right)^{2} \leq d \sum_{i=1}^{d} u_i^{2}$.

Hence, we get

$$
\left( \mathbb{E}_{\Theta} \left[ \left( \mathbb{E}_{X|\Theta} \left[ |g(X, \Theta)|^2 \mid \Theta = \theta \right] \right)^{\frac{2}{2}} \right] \right)^{\frac{2}{2}}
\leq \left( \mathbb{E}_{(X, \Theta)} \left[ g^2(X, \Theta) \right] + \mathbb{E}_{\Theta} \left[ \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \left[ \log \mu(\theta) \right] \right)^{2} \right)^{\frac{2}{2}}
\leq \left( (d \mathbb{E}_{\Theta} [ \text{Tr}(I(X(\Theta))) ] + d \text{Tr}(I(\mu)))^{\frac{2}{2}} \right), \quad (102)
$$

where the last inequality follows from $\left( \sum_{i=1}^{d} u_i \right)^{2} \leq d \sum_{i=1}^{d} u_i^{2}$, for non-negative $\{u_i\}_{i=1}^{d}$.

Summarizing, combining (88), (93) and (102) leads us to the desired lower bound

$$
\mathbb{E}_{(X, \Theta)} \left[ \left| \frac{\partial}{\partial \theta_i} \psi_i(\theta) \right| \right]^{p}
\geq \frac{d}{\text{Tr}(I(\mu))} \left[ (d \mathbb{E}_{\Theta} [ \text{Tr}(I(X(\Theta))) ] + d \text{Tr}(I(\mu)))^{\frac{2}{2}} \right],
$$

where $2$ Case $1 < p < 2$: Let $q \in \mathbb{R}$ such that $\frac{1}{2} + \frac{1}{q} = 1$, i.e., $q = p/p-1 \leq 2$. Also, consider the following two functions $g(\cdot)$ and $h(\cdot)$ defined, for $x \in \mathcal{X}$ and $\theta = [\theta_1, \ldots, \theta_d] \in \Theta$, as

$$
g(x, \theta) = \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \left[ \log f(x|\theta) \mu(\theta) \right] \quad (103a)
$$

$$
h(x, \theta) = \psi_i(\theta(x)) - \psi_i(\theta). \quad (103b)
$$

For convenience, for $i = 1, \ldots, d$ we will denote the $i^{th}$ component of $h(x, \theta)$ as $h_i(x, \theta)$, i.e.,

$$
h_i(x, \theta) = \psi_i(\theta(x)) - \psi_i(\theta) = (h(x, \theta))_i \quad (104)$$
Using the definition of the $p$-norm, the average estimation error can be lower bounded as
\[
E(X, \Theta) \left[ \left\| \psi(\theta(X)) - \psi(\theta) \right\|_p^p \right] = \sum_{i=1}^{d} E(X, \Theta) \left[ |h_i(X, \Theta)|^p \right]. \tag{105}
\]
The RHS of (105) can be lower bounded as follows. First, note that applying Hölder’s inequality for expectations yields
\[
E(X, \Theta) \left[ |h_i(X, \Theta)|g(X, \Theta) \right] \leq \left( E(X, \Theta) \left[ |h_i(X, \Theta)|^p \right] \right)^{\frac{1}{p}} \cdot \left( E(X, \Theta) \left[ |g(X, \Theta)|^{q'} \right] \right)^{\frac{1}{q'}}. \tag{106}
\]
Using the fact that the prior measure $\mu$ converges to zero at the endpoints of $\Theta$ and partial integration, it is easy to see that
\[
\int_{\Omega_1} h_i(x, \theta) \frac{\partial}{\partial \theta_i} \left[ f(x(\theta)) \mu_i(\theta_i) \right] \, d\theta_i = h_i(x, \theta) f(x(\theta)) \mu_i(\theta_i) \, d\theta_i - \int_{\Omega_1} h_i(x, \theta) f(x(\theta)) \mu_i(\theta_i) \, d\theta_i.
\]
Integration in (107), we get for $i = 1, \ldots, d$, that
\[
\int_{\Omega_1} h_i(x, \theta) \frac{\partial}{\partial \theta_i} \left[ f(x(\theta)) \mu_i(\theta_i) \right] \, d\theta_i \, dx = -E(X, \Theta) \left[ \frac{\partial}{\partial \theta_i} \left[ h_i(X, \Theta) \right] \right]. \tag{108}
\]
Thus, with some algebraic manipulations,
\[
E(X, \Theta) \left[ |h_i(X, \Theta)|g(X, \Theta) \right] = \sum_{i=1}^{d} E(X, \Theta) \left[ \left\{ \frac{\partial}{\partial \theta_i} \left[ h_i(X, \Theta) \right] \right\} \right] \leq \sum_{i=1}^{d} \left( E(X, \Theta) \left[ |h_i(X, \Theta)|^p \right] \right)^{\frac{1}{p}} \cdot \left( E(X, \Theta) \left[ |g(X, \Theta)|^{q'} \right] \right)^{\frac{1}{q'}}. \tag{109}
\]
Combining (105), (106), and (109), we get
\[
E(X, \Theta) \left[ \left\| \psi(\theta(X)) - \psi(\theta) \right\|_p^p \right] \geq d \sum_{i=1}^{d} \left( \frac{\partial}{\partial \theta_i} \left[ h_i(X, \Theta) \right] \right) \cdot \left( E(X, \Theta) \left[ |g(X, \Theta)|^p \right] \right)^{-\frac{p}{q'p}}. \tag{111}
\]
We now upper bound the second expectation of the RHS of (111), as follows. For convenience, let
\[
l(x, \theta) = \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \left[ \log f(x|\theta) \right]. \tag{112}
\]
It is easy to see that for all $\theta$, we have
\[
E_X[|l(X, \Theta)| \Theta = \theta] = 0 \tag{113}
\]
which follows by the regularity condition
\[
E_X[\frac{\partial}{\partial \theta_i} \log f(x|\theta)] = 0 \text{ for all } \theta \in \Theta.
\]
Thus, with some algebraic manipulations,
\[
\left( E(X, \Theta) \left[ \log(\theta(X)) \right] \right)^{\frac{1}{q'}} \leq \sum_{i=1}^{d} \left( E(X, \Theta) \left[ \left\| \frac{\partial}{\partial \theta_i} \log f(x|\theta) \right\|^p \right] \right)^{\frac{1}{p'}}. \tag{115}
\]
Combining (106) and (115), we have
\[
l(x, \theta) = \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \left[ \log f(x|\theta) \right]. \tag{116}
\]
Continuing from (116), the first term of its RHS can be upper bounded as
\[
\left( E(X, \Theta) \left[ \left\| \psi(\theta(X)) - \psi(\theta) \right\|_p^p \right] \right)^{\frac{1}{p'}} \leq \sum_{i=1}^{d} \left( E(X, \Theta) \left[ \left\| \frac{\partial}{\partial \theta_i} \log f(x|\theta) \right\|^p \right] \right)^{\frac{1}{p'}}. \tag{117}
\]
Thus, for all $i$ the inequality $f_i$, (1) Case $E_\theta - \theta = \theta$, $m_i$, $\ldots, \theta_i \in \Theta$, (d) holds by defining, for $i = 1, \ldots, d$ and $\theta \in \Theta$,

$$u_i(\theta) = \left( \text{E}_\theta \left[ \left[ \frac{\partial}{\partial \theta_i} \log f(X) \right]_{\theta = \theta} \right] \right)^{p-1}.$$

(e) holds by using the inequality $\sum_{i=1}^d u_i^{1/p} \leq \left( \sum_{i=1}^d u_i \right)^{1/p}$ for non-negative $(u_1, \ldots, u_d)$ and $p < 2$; and (f) holds by substituting using (121).

Substituting (117) in (116), we obtain

$$\left( \text{E}_\theta \left[ \left| g(X, \Theta) / \theta \right| \right] \right)^{1/p} \leq d^{1/p} \left( \text{E}_\theta \left[ \left| \Omega(X) \right| \right] \right)^{1/2} + d^{1-1/p} \left( \text{E}_\theta \left[ \left| \Omega \right| \right] \right)^{1/2}.$$ (119)

Substituting (119) in (111) produces the desired lower bound

$$\text{E}_\theta \left[ \left| \psi(\hat{\theta}(X)) - \psi(\theta) \right| \right] \leq \left( \text{E}_\theta \left[ \left| \Omega \right| \right] \right)^{1/2} + \left( \text{E}_\theta \left[ \left| \Omega \right| \right] \right)^{1/2}.$$ (120)

D. Proof of Theorem 2

1) Case $p \geq 2$.

Let $q \in \mathbb{R}$ such that $\frac{1}{p} + \frac{1}{q} = 1$, i.e., $q = p/(p-1)$. Also, consider the following two functions $g(\cdot)$ and $h(\cdot)$ defined for $X, \Theta = \{\theta_1, \ldots, \theta_d\} \subset\subset \Theta$ and a specific quantization messages tuple $m^{(n)} = (m_1, \ldots, m_n) \in \{1, 2^n\}$ as

$$g(m^{(n)}, \theta) = \sum_{i=1}^d \frac{\partial}{\partial \theta_i} \left[ \log \left( p(m^{(n)}(\theta)) \mu(\theta) \right) \right],$$ (120a)

$$h(m^{(n)}, \theta) = \hat{\theta}(m^{(n)}) - \theta.$$ (120b)

where in (120a) the quantization messages joint probability is $p(m^{(n)}(\theta)) = \prod_{j=1}^n \mathcal{P}(m_j | \theta_j)$. For convenience, for $i = 1, \ldots, d$ we will denote the $i^{th}$ component of $h(m^{(n)}, \theta)$ as $h_i(m^{(n)}, \theta)$, i.e.,

$$h_i(m^{(n)}, \theta) = \hat{\theta}_i(m^{(n)}) - \theta_i = \left( h(m^{(n)}, \theta) \right)_i.$$ (121)

Using the fact that the prior measure $\mu$ converges to zero at the boundaries of $\Theta$, it is easy to see that

$$\sum_{m^{(n)}} \int_{\Theta} h_i(m^{(n)}, \theta) \frac{\partial}{\partial \theta_i} \left[ p(m^{(n)}(\theta)) \mu(\theta) \right] d\theta_i = 1.$$ (122)

By partial integration and (122), we get for $i = 1, \ldots, d$, that

$$\sum_{m^{(n)}} \int_{\Theta} h_i(m^{(n)}, \theta) \frac{\partial}{\partial \theta_i} \left[ p(m^{(n)}(\theta)) \mu(\theta) \right] d\theta_i = 1.$$ (123)

Thus, for all $i = 1, \ldots, d$, we have

$$d \leq \text{E}_\theta \left[ \left| h_i(M^{(n)}, \Theta) g(M^{(n)}, \Theta) \right| \right].$$ (124)

Note that we have

$$\text{E} \left[ \left| h_i(M^{(n)}, \Theta) g(M^{(n)}, \Theta) \right| \right]$$

$$= \text{E}_\theta \left[ \left| h_i(M^{(n)}, \Theta) g(M^{(n)}, \Theta) \right| \right]$$

$$\leq \text{E}_\theta \left[ \left( \left| h_i(M^{(n)}, \Theta) g(M^{(n)}, \Theta) \right| \right)^2 \right]^{1/2}$$

$$\times \left( \text{E}_\theta \left[ \left( g(M^{(n)}, \Theta) \right)^2 \right] \right)^{1/2}$$

$$\leq \text{E}_\theta \left[ \left( h_i(M^{(n)}, \Theta) \right)^2 \right]^{1/2} \times \left( \text{E}_\theta \left[ \left( g(M^{(n)}, \Theta) \right)^2 \right] \right)^{1/2},$$ (125)

where (a) follows by application of Hölder’s inequality for every $\theta \in \Theta$ to the conditional expectation $\text{E}_\theta \left[ \left( h_i(M^{(n)}, \Theta) \right)^2 \right]$; and (b) follows by application of Hölder’s inequality to the expectation $\text{E}_\theta \left[ \left( h_i(M^{(n)}, \Theta) \right)^2 \right]$ since $p > 1$, $q > 1$ and are such that $\frac{1}{p} + \frac{1}{q} = 1$.

The first element of the right-hand side of (125) produces the desired risk

$$\sup_{\theta \in \Theta} \text{E}_\theta \left[ \left( \hat{\theta}(M^{(n)}) - \theta \right)^p \right]$$

$$\geq \text{E}_\theta \left[ \left( \hat{\theta}(M^{(n)}) - \theta \right)^p \right]$$

$$\geq \text{E}_\theta \left[ \left( \hat{\theta}(M^{(n)}) - \theta \right)^p \right]$$

$$\geq \sum_{i=1}^d \text{E}_\theta \left[ \left( h_i(M^{(n)}, \Theta) \right)^2 \right] \left( \Theta \right)^{\frac{1}{q}}.$$ (127)

where in $(a)$ the supremum bounds the expectation, $(b)$ follows by the definition of the $p$-norm and $(c)$ by replacing $E_{M^{(n)} \Theta}$ with $E_{M \Theta}$ and $E_{M^{(n)} \Theta}$ with and Jensen’s inequality for expectations for convex functions $x \mapsto x^\gamma$, for $2 < p$.

Combining (124), (125) and (127), we get

$$\sup_{\theta \in \Theta} \text{E}_\theta \left[ \left( \hat{\theta}(M^{(n)}) - \theta \right)^p \right]$$

$$d^{p+1} \left( \text{E}_\theta \left[ \left( \text{E}_{M^{(n)}} \left[ \left( \left| g(M^{(n)}, \Theta) \right| \right)^2 \right] \right)^{1/2} \right] \right)^{\frac{1}{q}}.$$ (128)

We now move on to the last step of the proof, to upper bound the expectation of the RHS of (128). For convenience, let

$$l(\Theta) = \sum_{i=1}^d \frac{\partial}{\partial \theta_i} \left[ \log p(m_j | \Theta) \right] \text{ with } E_{M \Theta} \left[ l(M_j, \Theta) \right] = 0.$$ (129)

Note that $l(m_j, \Theta)$ is the sum of the elements of the score function associated with $M_j$. We expand the square and cancel the product of the two elements, due to the property that $E_{M_j \Theta} \left[ l(M_j, \Theta) \right] = 0,$
to arrive at the trace of the Fisher information matrix of $M_j$ and that of the prior, respectively, as follows

$$
\left( E_{M^{(n)}|\Theta} \left[ \left( g(M^{(n)}|\Theta, \Theta) \right)^2 | \Theta \right] \right)^{\frac{1}{2}} = 
\left( E_{M^{(n)}|\Theta} \left[ \left( \sum_{j=1}^{n} l(M_j, \Theta) \right)^2 | \Theta \right] \right)^{\frac{1}{2}}
+ \left( \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2 \right)^{\frac{1}{2}}
$$

(130)

which holds by

$$
E_{M^{(n)}|\Theta} \left[ \left( \sum_{j=1}^{n} l(M_j, \Theta) \right)^2 \right] \left( \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2 = 0.
$$

Further, by Jensen's inequality for expectations for concave functions $x \mapsto x^\frac{1}{2}$, for $q < 2$, we have

$$
\left( E_\Theta \left[ \left( E_{M^{(n)}|\Theta} \left[ \left( g(M^{(n)}|\Theta, \Theta) \right)^2 | \Theta \right] \right)^{\frac{1}{2}} \right] \right)^{\frac{1}{2}} \leq
\left( E_{(M^{(n)}, \Theta)} \left[ \left( \sum_{j=1}^{n} l(M_j, \Theta) \right)^2 \right] \right)^{\frac{1}{2}}
+ \left( E_{(M^{(n)}, \Theta)} \left[ \left( \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2 \right] \right)^{\frac{1}{2}}
+ \left( \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2 \right)^{\frac{1}{2}}
$$

(131)

For $l(M_j, \Theta)$ with $E_{M_j|\Theta} [l(M_j, \Theta)] = 0$ and independent, by the Marcinkiewicz-Zygmund inequality in the form of (2) of [19], there exists a constant $B_2 = 1$ [19], such that

$$
E_{M^{(n)}|\Theta} \left[ \sum_{j=1}^{n} l(M_j, \Theta) \right]^2 =
B_2 E_{M^{(n)}|\Theta} \left[ \sum_{j=1}^{n} l^2(M_j, \Theta) \right] \leq
\left( \sum_{j=1}^{n} E_{M^{(n)}|\Theta} \left[ \left( \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \log p(M_j|\Theta) \right)^2 \right] | \Theta \right)
\leq
\sum_{j=1}^{n} E_{M^{(n)}|\Theta} \left[ \left( \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \log p(M_j|\Theta) \right)^2 \right]
\leq
\sum_{j=1}^{n} \text{Tr}(I_{M_j}(\Theta))
$$

(132)

where, by the independence of $M_j$, the expectation of each element of the summation is identical and this leads to the term $n$ in (a), which also follows from the definition of $l(M_j, \Theta)$, (b) is given by the inequality $(\sum_{i=1}^{d} x_i)^2 \leq d \sum_{i=1}^{d} x_i^2$, $x_i > 0$, required in order to pass the summation inside the expectation and obtain the trace of the Fisher information matrix for $M_j$.

Substituting (132) in (131), we obtain

$$
\left( E_\Theta \left[ \left( E_{M^{(n)}|\Theta} \left[ \left( g(M^{(n)}|\Theta, \Theta) \right)^2 | \Theta \right] \right)^{\frac{1}{2}} \right] \right)^{\frac{1}{2}} \leq
\left( d E_\Theta \left[ \sum_{j=1}^{n} \text{Tr}(I_{M_j}(\Theta)) \right] + d \text{Tr}(I(\mu)) \right)^{\frac{1}{2}}.
$$

(133)

Substituting (133) in (128), produces the desired lower bound

$$
\sup_{\theta \in \Theta} E_{M^{(n)}|\Theta} \left[ \left| \hat{\theta}(M^{(n)}) - \theta \right|^p | \Theta \right] \geq
\left( \sum_{j=1}^{n} E_\Theta \left[ \text{Tr}(I_{M_j}(\Theta)) \right] + \text{Tr}(I(\mu)) \right)^{-\frac{p}{2}}.
$$

(134)

Remark 3. For a random variable with bounded support, $\theta \in [-B, B]$, the prior distribution $\mu(\theta)$ that minimizes the Fisher information is the raised cosine distribution. That is, for $\theta \in [-B, B]$

$$
\mu(\theta) = \frac{1}{B} \cos^2 \left( \frac{\pi \theta}{2B} \right),
I(\mu) = \frac{\pi^2}{B^2} \Omega(p)(\mu) =
\sum_{i=1}^{d} l(p)(\mu_i) = d \left( E_{\Theta_i} \left[ \left( \frac{\partial}{\partial \Theta_i} \log \mu_i(\Theta_i) \right)^p \right] \right)^{p-1}
\leq
\frac{\pi d}{2} \left( \frac{2}{B} \right)^p \left[ B \left( \frac{2p - 1}{2p} \frac{1}{2p - 2} \right) \right]^{p-1},
$$

(135)

where $I(\mu_i)$ represents the Fisher information associated to the prior. The condition $p > 1.5$ is required to ensure the existence of the Beta function $B(.)$.

E. Proof of Theorem 2

1) Case $p \geq 2$: By Theorem 2, if $p \geq 2$, we have that

$$
\sup_{\theta \in \Theta} E_{M^{(n)}|\Theta} \left[ \left| \hat{\theta}(M^{(n)}) - \theta \right|^p | \Theta \right] \geq
\left( \sum_{j=1}^{n} E_{M^{(n)}|\Theta} \left[ \text{Tr}(I_{M_j}(\Theta)) \right] + \text{Tr}(I(\mu)) \right)^{-\frac{p}{2}}.
$$

We need to compute an upper bound on $\text{Tr}(I_{M_j}(\Theta))$. If $p \geq 2$, then, Theorem E gives us that, for some $r \geq 1$, the upper bound holds

$$
\text{Tr}(I_{M_j}(\Theta)) \leq \min \left\{ \text{Tr}(I_{X}(\Theta)), 4I_0 k^2 \right\}.
$$

Then, also by Remark 3 we obtain

$$
\sup_{\theta \in \Theta} E_{M^{(n)}|\Theta} \left[ \left| \hat{\theta}(M^{(n)}) - \theta \right|^p | \Theta \right] \geq
\left( 4I_0 k^2 n + \frac{d \pi^2}{B^2} \right)^{-\frac{p}{2}}.
$$

(136)
2) Case $1 < p < 2$: By Theorem 3 if $1 < p < 2$, we have that

$$
\sup_{\theta \in \Theta} E_{M^{(n)}}[\left\| \hat{\theta}(M^{(n)}) - \theta \right\|_p^p] \geq d^p \left[ \frac{d^{2p-2}}{p-1} \left( \Omega_p(\mu) \right)^\frac{1}{2p} \right]^{-p} + \frac{1}{p-1} \left( \sum_{j=1}^n \left( E_{\Theta} \left( \left( \Omega_{M_j}(\theta) \right)_r^{\frac{1}{p}-1} \right) \right)^{2(p-1)} \right)^{\frac{1}{2}} \left( \frac{2}{p} \right)^{\frac{1}{2}} \left( \frac{2}{p} \right)^{-p}.
$$

We need to compute an upper bound on $\Omega_{M_j}(\theta)$. If $1 < p < 2$, then, Theorem 5 gives us that, for some $r \geq \frac{p}{2n-2}$, the upper bound holds

$$
\Omega_{M_j}(\theta) \leq \min \{ \Omega_{M_j}(\theta), d^{\frac{2p-2}{p-1}} I_0^\frac{1}{2} (2k)^{-2p} k^\frac{\mu}{2} \}.
$$

Then, also by Remark 3, we get

$$
I_{(p)}(\mu_i) = \frac{\pi}{2} \left( \frac{2}{B} \right)^p B \left( \frac{2p-1}{2p-2} \right) \left( \frac{2p-3}{2p-2} \right)^{p-1} \left( \frac{2p-1}{2p-2} \right)^{p-1} \left( \frac{2p-3}{2p-2} \right)^{p-1} = \left( \frac{2}{B} \right)^p \left( \frac{2p-1}{2p-2} \right)^{p-1} \left( \frac{2p-3}{2p-2} \right)^{p-1}.
$$

F. Proof of Corollary 2

Using the expression of the score function,

$$
S_\theta(X) = \frac{\partial}{\partial \theta} \log f(x|\theta) = \frac{1}{\sigma^2}(x - \theta),
$$

we compute

$$
\Omega_{X}(\theta) = \sum_{i=1}^d \left( \mathbb{E}_{\Theta}[|S_{\theta}(X)|_{\frac{1}{p} - 1}] \right)^{p-1} = \left( \frac{1}{\sigma^2} \right)^p \sum_{i=1}^d \left( \mathbb{E}_{X|\theta}[|X_i - \theta_i|_{\frac{1}{p} - 1}] \right)^{p-1} = \left( \frac{1}{\sigma^2} \right)^p \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^{p-1} \times \sum_{i=1}^d \left( \int_{x_i} |x_i - \theta_i|_{\frac{1}{p} - 1} \exp \left[ \frac{-(x_i - \theta_i)^2}{2\sigma^2} \right] dx_i \right)^{p-1}.
$$

Using the gamma function, we compute the above integral as

$$
\int_{x_i} |x_i - \theta_i|_{\frac{1}{p} - 1} \exp \left[ \frac{-(x_i - \theta_i)^2}{2\sigma^2} \right] dx_i = \frac{2^{\frac{1}{2p-2}} \sigma^{\frac{2}{p-1}}}{p-1} \Gamma \left( \frac{1}{2p-2} \right).
$$

Then, we obtain further

$$
\Omega_{X}(\theta) = \frac{d^{\frac{2}{p}}}{\sigma^\frac{2}{p}} \left( \frac{\Gamma \left( \frac{1}{2p-2} \right)}{(p-1)\sqrt{2\pi\sigma^2}} \right)^{p-1}.
$$

1) Case $p \geq 2$: By Theorem 3 if $p \geq 2$, we have that

$$
\sup_{\theta \in \Theta} E_{M^{(n)}}[\left\| \hat{\theta}(M^{(n)}) - \theta \right\|_p^p] \geq d^p \left[ \frac{d^{2p-2}}{p-1} \left( \Omega_p(\mu) \right)^\frac{1}{2p} \right]^{-p} + \frac{1}{p-1} \left( \sum_{j=1}^n \left( E_{\Theta} \left( \left( \Omega_{M_j}(\theta) \right)_r^{\frac{1}{p}-1} \right) \right)^{2(p-1)} \right)^{\frac{1}{2}} \left( \frac{2}{p} \right)^{\frac{1}{2}} \left( \frac{2}{p} \right)^{-p}.
$$

We need to compute an upper bound on $\Omega_{M_j}(\theta)$. If $p \geq 2$, then, Theorem 5 gives us that, for some $r \geq 1$, the upper bound holds

$$
\Omega_{M_j}(\theta) \leq \min \{ \Omega_{M_j}(\theta), d^{\frac{2p-2}{p-1}} I_0^\frac{1}{2} (2k)^{-2p} k^\frac{\mu}{2} \}.
$$

We move on to compute the value of $I_0$. That is, using the same approach as in Corollary 1 of [2], for $r = 2 \geq 1$, we obtain $I_0 = \frac{8}{3\sigma^2}$. Then, by Remark 3, we get

$$
\sup_{\theta \in \Theta} E_{M^{(n)}}[\left\| \hat{\theta}(M^{(n)}) - \theta \right\|_p^p] \geq d^{(1+\frac{2}{p})} \left( \frac{n d}{\sigma^2} + \frac{d \pi^2 \mu^2}{B^2} \right)^{-\frac{1}{2}} \left( \frac{2 \pi \sigma^2}{3\sigma^2} \right)^{-\frac{1}{2}}.
$$

For $\pi^2 \sigma^2 d \leq n B^2 \min \{k, d\}$, we can ignore the prior to obtain

$$
\sup_{\theta \in \Theta} E_{M^{(n)}}[\left\| \hat{\theta}(M^{(n)}) - \theta \right\|_p^p] \geq d^{(1+\frac{p}{2})} \left( \frac{n d}{\sigma^2} + \frac{d \pi^2 \mu^2}{B^2} \right)^{-\frac{1}{2}} \left( \frac{2 \pi \sigma^2}{3\sigma^2} \right)^{-\frac{1}{2}}.
$$

2) Case $1 < p < 2$: By Theorem 3 if $1 < p < 2$, we have that

$$
\sup_{\theta \in \Theta} E_{M^{(n)}}[\left\| \hat{\theta}(M^{(n)}) - \theta \right\|_p^p] \geq d^p \left[ \frac{d^{2p-2}}{p-1} \left( \Omega_p(\mu) \right)^\frac{1}{2p} \right]^{-p} + \frac{1}{p-1} \left( \sum_{j=1}^n \left( E_{\Theta} \left( \left( \Omega_{M_j}(\theta) \right)_r^{\frac{1}{p}-1} \right) \right)^{2(p-1)} \right)^{\frac{1}{2}} \left( \frac{2}{p} \right)^{\frac{1}{2}} \left( \frac{2}{p} \right)^{-p}.
$$

We need to compute an upper bound on $\Omega_{M_j}(\theta)$. If $1 < p < 2$, then, Theorem 5 gives us that, for some $r \geq \frac{p}{2n-2}$, the upper bound holds

$$
\Omega_{M_j}(\theta) \leq \min \{ \Omega_{X}(\theta), d^{\frac{2p-2}{p-1}} I_0^\frac{1}{2} (2k)^{-2p} k^\frac{\mu}{2} \}.
$$

We move on to compute the value of $I_0$. That is, using the same approach as in Corollary 1 of [2], for $r = 2 \geq \frac{p}{2n-2}$ for $p > 1$, we obtain $I_0 = \frac{8}{3\sigma^2}$. Then, we obtain

$$
\Omega_{M_j}(\theta) \leq \min \left\{ \frac{d^{\frac{1}{2}}}{\sigma} \left[ \frac{(1-\frac{2}{p-2})}{(p-1)\sqrt{2\pi\sigma^2}} \right]^{p-1}, \frac{2p+k(2-p)}{d^{\frac{1}{2}}} \left( \frac{8k}{3\sigma^2} \right)^{\frac{1}{2}} \right\}.
$$
Then, also by Remark 3, $v_p > 1.5$, which is required for the Beta function $B(\cdot, \cdot)$ to exist, we obtain

$$\Omega(p)(\mu) = \frac{\pi d}{2} \left( \frac{2}{B} \right)^p B \left( \frac{2p-1}{2}, \frac{2p-3}{2} \right) p^{-1}$$

and

$$\sup_{\Theta \in \Theta} E_{M(n)} \left[ \left\| \hat{\theta}(M(n)) - \theta \right\|_p^p \right] \geq d^p \times \max \left\{ \left( \frac{\sqrt{\pi}}{p-1} \left( \frac{2}{B} \right)^{\frac{1}{2}} \left[ \left( \frac{2p-1}{2} \right) \left( \frac{2p-2}{2} \right) \right]^{-\frac{1}{2}} + p, \left( \frac{\sqrt{\pi}}{p-1} \left( \frac{2}{B} \right)^{\frac{1}{2}} \left[ \left( \frac{2p-1}{2} \right) \left( \frac{2p-2}{2} \right) \right]^{-\frac{1}{2}} + p \right)^{\frac{1}{2}} \right\} \right.$$

$$+ d^{p-1} \left( \frac{\pi}{2} \right)^{\frac{1}{2}} \left( \frac{2}{B} \right)^{\frac{1}{2}} B \left( \frac{2p-1}{2}, \frac{2p-3}{2} \right) p^{-1} \nu$$

$$+ d^{p-1} \left( \frac{\pi}{2} \right)^{\frac{1}{2}} \left( \frac{2}{B} \right)^{\frac{1}{2}} B \left( \frac{2p-1}{2}, \frac{2p-3}{2} \right) p^{-1} \nu.$$

G. Proof of Theorem 4

From the definition of the Wasserstein distance, we have

$$E_{M(n)} \left[ W_p(f(x|\hat{\theta}(M(n))), f(x|\theta)) \right] = E_{M(n)} \left[ E_{Z,Y} \sim \mu^* \left[ d^p(Z,Y) \right] \right]$$

$$= E_{M(n)} \left[ \int_{z,y} \int_{z,y} d^p(z,y) \mu^*(z,y) dz dy \right]$$

If $d(z,y) = \|z - y\|_p$, then

$$E_{M(n)} \left[ W_p(f(x|\hat{\theta}(M(n))), f(x|\theta)) \right] = E_{M(n)} \left[ E_{z,y} \sim \mu^* \left[ d^p(z,y) \right] \right]$$

$$= E_{M(n)} \left[ \int_{z,y} \left( \int_{z,y} \left( \sum_{i=1}^{d} |z_i - y_i|^p \right) \mu^*(z,y) dz dy \right) \right]$$

$$= E_{M(n)} \left[ \left( \sum_{i=1}^{d} |Z_i - Y_i|^p \right) \right]$$

$$\geq E_{M(n)} \left[ \left( \sum_{i=1}^{d} E_{z,y} \sim \mu^* \left[ |Z_i - Y_i| \right] \right) \right]$$

$$= E_{M(n)} \left[ \left( \sum_{i=1}^{d} E_{z,y} \sim \mu^* \left[ |Z_i - Y_i| \right] \right) \right]$$

$$= E_{M(n)} \left[ \left( \sum_{i=1}^{d} \left[ E_{Z,Y} \sim f(z|\theta(M(n))) \left| |Z_i - Y_i| \right| \right] \right) \right]$$

where (a) follows from Jensen’s inequality for convex functions of expectations, $E[|X|^p] \geq (E[|X|])^p$, $p > 1$ and (b) is given by $E[|X|] \geq |E[X]|$.

1) Case $1 < p < 2$: Let $q \in \mathbb{R}$ such that $\frac{1}{2} + \frac{1}{p} = 1$, i.e., $q = \frac{p}{p-1}$. Also, consider the following two functions $g(\cdot)$ and $h(\cdot)$ defined, for $x \in X$, $\theta = [\theta_1, \ldots, \theta_d] \in \Theta$ and a specific quantization messages tuple $m^{(n)} = (m_1, \ldots, m_n) \in [1, 2^k]^n$ as

$$g(m^{(n)}|\theta) = \sum_{i=1}^{d} \frac{\partial}{\partial \theta_i} \left[ \log \left( p(m^{(n)}|\theta)\mu(\theta) \right) \right]$$

$$h(m^{(n)}|\theta) = E_{Z \sim f(z|\theta(M(n)))} \left| |Z| - E_{Y \sim f(y|\theta)} \right| Y.$$
Thus, with some algebraic manipulations,
\[
E_{(M^{(n)}, \theta)} \left[ h_j(M^{(n)}, \theta) g(M^{(n)}, \theta) \right] \\
= \sum_{i=1}^{d} E_{\Theta_i} \left[ \cdots E_{\Theta_d} \left[ -E_{(M^{(n)}, \Theta)} \left[ \frac{\partial}{\partial \Theta_j} \left[ h_j(M^{(n)}, \theta) \right] \right] \right] \right] \\
= -\sum_{i=1}^{d} E_{(M^{(n)}, \Theta)} \left[ \frac{\partial}{\partial \Theta_j} \left[ h_j(M^{(n)}, \theta) \right] \right] \\
= \sum_{i=1}^{d} E_{\Theta_i} \left[ \frac{\partial}{\partial \Theta_j} \left[ E_{Y \sim f(y|\theta)} [Y_j] \right] \right] \\
\text{and } |E[X]| \leq |E[|X|]| \text{ lower bounds the left-hand side of (146) as}
\left| E_{(M^{(n)}, \Theta)} \left[ h_j(M^{(n)}, \theta) g(M^{(n)}, \theta) \right] \right| \\
\leq \sum_{i=1}^{d} E_{\Theta_i} \left[ \frac{\partial}{\partial \Theta_j} \left[ E_{Y \sim f(y|\theta)} [Y_j] \right] \right]. \tag{150}
\]

Combining (145), (146), (147) and (150), we get
\[
E_{(M^{(n)}, \Theta)} \left[ W^p_p(f(x|\hat{\theta}(M^{(n)}), f(x|\theta)) \right] \geq \left( \sum_{j=1}^{d} \sum_{i=1}^{d} E_{\Theta_i} \left[ \frac{\partial}{\partial \Theta_j} \left[ E_{Y \sim f(y|\theta)} [Y_j] \right] \right] \right)^p \times \left( E_{(M^{(n)}, \Theta)} \left[ g(M^{(n)}, \theta) \right] \right)^{-\frac{p}{q}} \tag{151} \]

We now upper bound the second expectation term of the RHS of (151). For convenience, let for \( j = 1, \ldots, 2^k \)
\[
I(m_j, \theta) = \sum_{i=1}^{d} \frac{\partial}{\partial \Theta_i} \left[ \log p(m_j|\theta) \right]. \tag{152}
\]
It is easy to see that for all \( \theta \), we have
\[
E_{M_j|\theta} \left[ I(M_j, \theta) \right] = 0. \tag{153}
\]
Then, we have
\[
\left( E_{(M^{(n)}, \Theta)} \left[ g(M^{(n)}, \theta) \right] \right)^{-\frac{p}{q}} \leq \left( E_{(M^{(n)}, \Theta)} \left[ \sum_{j=1}^{n} I(M_j, \theta) \right] \right)^{\frac{q}{p}} + d^{\frac{q}{p}} \left( \Omega^p(\mu) \right)^{\frac{1}{p}}, \tag{154} \]
where the inequality holds by a double application of Minkowski's inequality: first for expectations using that for all \( Z \) and \( T \) we have \( |E[|Z + T|^q]| \leq |E[|Z|^q]|^\frac{1}{q} + |E[|T|^q]|^\frac{1}{q} \) and then that
\[
|E[\sum_{i=1}^{d} Z_i]|^\frac{1}{q} \leq \sum_{i=1}^{d} |E[Z_i]|^\frac{1}{q}, \forall q > 1 \text{ and } \sum_{i=1}^{d} u_i \leq d^{\frac{1}{q}} \max_{i=1}^{d} u_i, \forall u_i > 0, p > 1.
\]
Next, since the quantities \( \{I(M_j, \theta)\}_j \) are independent and satisfy that \( E_{M_j|\theta} \left[ I(M_j, \theta) \right] = 0 \) for all \( j = 1, \ldots, 2^k \), the application of Marcinkiewicz-Zygmund inequality [18, 19] yields
\[
E_{M^{(n)}, \Theta} \left[ \sum_{j=1}^{n} I(M_j, \theta) \right] \leq B_q \exp \left( \frac{1}{2} \left( \sum_{j=1}^{n} I^2(M_j, \theta) \right)^{\frac{q}{2}} \right) \tag{155} \]
where \( B_q = 1/(q - 1)^q > 0. \)
Continuing from (155), we get
\[
\left( E_{(M^{(n)}, \Theta)} \left[ \sum_{j=1}^{n} I(M_j, \theta) \right] \right)^{\frac{q}{2}} \leq \left( (p - 1)^2 \sum_{j=1}^{n} E_{(M^{(n)}, \Theta)} \left[ \log p(M_j|\Theta) \right] \right)^{\frac{q}{2}} \tag{156a} \]
\[
\leq d^\frac{q}{2} (p - 1)^2 \sum_{j=1}^{n} E_{(M^{(n)}, \Theta)} \left[ \log p(M_j|\Theta) \right]^{\frac{q}{2}} \tag{156b} \]
\[
\leq d^\frac{q}{2} (p - 1)^2 \sum_{j=1}^{n} \left( E_{(M^{(n)}, \Theta)} \left[ \log p(M_j|\Theta) \right] \right)^{\frac{q}{2}} \tag{156c} \]
where (a) follows by using the quantization messages are independent, substituting \( q = p/(p - 1) \) and applying Minkowski's inequality \( |E[\sum_{i=1}^{d} Z_i]|^\frac{1}{q} \leq \sum_{i=1}^{d} |E[Z_i]|^\frac{1}{q} \) since \( q = p > (p - 1)/2 \); (b) follows by substituting using (152) and using that \( \sum_{i=1}^{d} u_i \leq d^{1-q} \sum_{i=1}^{d} u_i > 0 \) and (c) holds by (13).

Finally, combining (156) with (154) and substituting in (151) yields the desired result
\[
\sup_{\Theta \in \Theta} E_{M^{(n)}, \Theta} \left[ W^p_p(f(x|\hat{\theta}(M^{(n)}), f(x|\theta)) \right] \geq \left( \sum_{j=1}^{d} \sum_{i=1}^{d} E_{\Theta_i} \left[ \frac{\partial}{\partial \Theta_j} \left[ E_{Y \sim f(y|\theta)} [Y_j] \right] \right] \right)^p \times \left( \sum_{j=1}^{n} \left( E_{\Theta_i} \left[ \log p(M_j|\Theta) \right] \right)^{\frac{1}{2}} \right)^{2(p-1)^{\frac{1}{p}}} \tag{157a} \]
\[
h(m^{(n)}, \theta) = E_{Z \sim f(x|\theta(M^{(n)}))} [Z] - E_{Y \sim f(y|\theta)} [Y]. \tag{157b} \]

where in (157a) the quantization messages joint probability is \( p(m^{(n)}|\theta) = \prod_{j=1}^{n} p_j(m_j|\theta) \). For convenience, for \( i = 1, \ldots, d \) we will denote the \( i \)-th component of \( h(m^{(n)}, \theta) \) as \( h_i(m^{(n)}, \theta) \), i.e.,
\[
h_i(m^{(n)}, \theta) = E_{Z \sim f(x|\theta(M^{(n)}))} [Z_i] - E_{Y \sim f(y|\theta)} [Y_i] = \left( h(m^{(n)}, \theta) \right)_i. \tag{158} \]
Note that we have
\[
\begin{align*}
\mathbb{E} \left[ h_i(M^{(n)}, \Theta) g(M^{(n)}, \Theta) \right] \\
= \mathbb{E}_{\Theta} \mathbb{E}_{X|\Theta} \left[ \left( \left( h_i(M^{(n)}, \Theta) g(M^{(n)}, \Theta) \right) \right) | \Theta = \theta \right] \\
\leq \mathbb{E}_{\Theta} \left( \mathbb{E}_{M^{(n)}|\Theta} \left[ \left( h_i(M^{(n)}, \Theta) g(M^{(n)}, \Theta) \right)^2 | \Theta = \theta \right] \right)^{\frac{1}{2}} \\
\times \left( \mathbb{E}_{\Theta} \left[ \left( g(M^{(n)}, \Theta) \right)^2 | \Theta = \theta \right] \right)^{\frac{1}{2}} \\
\leq \left( \mathbb{E}_{\Theta} \left[ \mathbb{E}_{M^{(n)}|\Theta} \left[ \left( h_i(M^{(n)}, \Theta) \right)^2 | \Theta = \theta \right] \right] \right)^{\frac{1}{2}} \\
\times \left( \mathbb{E}_{\Theta} \left[ \left( g(M^{(n)}, \Theta) \right)^2 | \Theta = \theta \right] \right)^{\frac{1}{2}}
\end{align*}
\] (159)

Thus, with some algebraic manipulations,
\[
\begin{align*}
&\mathbb{E}_{(M^{(n)}, \Theta)} \left[ h_j(M^{(n)}, \Theta) g(M^{(n)}, \Theta) \right] \\
= \sum_{i=1}^{d} \mathbb{E}_{\Theta_i} \left[ \cdots \mathbb{E}_{\Theta_d} \left[ -\frac{\partial}{\partial \Theta_j} \left( h_j(M^{(n)}, \Theta) \right) \right] \right] \\
= -\sum_{i=1}^{d} \mathbb{E}_{(M^{(n)}, \Theta)} \left[ \frac{\partial}{\partial \Theta_i} \left( h_j(M^{(n)}, \Theta) \right) \right] \\
= \sum_{i=1}^{d} \mathbb{E}_{(M^{(n)}, \Theta)} \left[ \frac{\partial}{\partial \Theta_i} \mathbb{E}_{Y \sim f(\Theta)} [Y_j] \right]
\end{align*}
\] (164)

and \( \mathbb{E}[X] \leq \mathbb{E}[|X|] \) lower bounds the left-hand side of (159) as
\[
\sum_{i=1}^{d} \mathbb{E}_{(M^{(n)}, \Theta)} \left[ \mathbb{E}_{Y \sim f(\Theta)} [Y_j] \right] \leq \mathbb{E}_{(M^{(n)}, \Theta)} \left[ h_j(X, \Theta) g(X, \Theta) \right].
\] (165)

Combining (159 b), (161) and (165), we get
\[
\sup_{\Theta \in \Theta} \mathbb{E}_{M^{(n)}|\Theta} \left[ W^2_{\theta}(f(x|\hat{\Theta}(M^{(n)})), f(x|\Theta)) | \Theta = \theta \right] \geq \sum_{j=1}^{d} \left( \sum_{i=1}^{d} \mathbb{E}_{(M^{(n)}, \Theta)} \left[ \frac{\partial}{\partial \Theta_j} \mathbb{E}_{Y \sim f(\Theta)} [Y_j] \right] \right)^{\frac{p}{2}} \\
\times \left( \mathbb{E}_{\Theta} \left[ \left( \mathbb{E}_{M^{(n)}|\Theta} \left[ \left( g(M^{(n)}, \Theta) \right)^2 | \Theta = \theta \right] \right) \right] \right)^{-\frac{p}{2}}
\] (166)

We now move on to the last step of the proof, to upper bound the second expectation of the RHS of (166). For convenience, let
\[
l(m, \Theta) = \sum_{j=1}^{d} \frac{\partial}{\partial \Theta_j} \left[ \log p(mj|\Theta) \right] \text{ with } E_{M|\Theta} [l(M, \Theta)] = 0.
\] (167)

Then, \( g(m^{(n)}, \Theta) = \sum_{j=1}^{d} l(m, \Theta) + \sum_{i=1}^{d} \frac{\partial}{\partial \Theta_i} \left[ \log \mu(\Theta) \right]. \)

Note that \( l(m, \Theta) \) is the sum of the elements of the score function associated with \( M_j \). We expand the square and cancel the product of the two elements, due to the property that \( E_{M_j|\Theta} [l(M, \Theta)] = 0 \), to arrive at the trace of the Fisher information matrix of \( M_j \) and that of the prior, respectively, as follows
\[
\mathbb{E}_{M^{(n)}|\Theta} \left[ \left( \mathbb{E}_{(M^{(n)}, \Theta)} \left[ \left( g(M^{(n)}, \Theta) \right)^2 | \Theta \right] \right)^{\frac{1}{2}} \\
+ \left( \sum_{j=1}^{d} \frac{\partial}{\partial \Theta_j} \left[ \log \mu(\Theta) \right] \right)^{\frac{1}{2}} \right] = 0.
\] (168)

which holds by
\[
\mathbb{E}_{M^{(n)}|\Theta} \left[ \left( \sum_{j=1}^{d} l(M_j, \Theta) \right) \right] + \left( \sum_{i=1}^{d} \frac{\partial}{\partial \Theta_i} \left[ \log \mu(\Theta) \right] \right) = 0.
\] (163)
Further, by Jensen’s inequality for expectations for concave functions $x \mapsto x^\frac{1}{2}$, for $q < 2$, we have

$$
\left( \mathbb{E}_\Theta \left[ \left( \mathbb{E}_{M^{(n)}|\Theta} \left[ \left| g(M^{(n)}, \Theta) \right|^2 \mid \Theta \right] \right)^\frac{q}{2} \right] \right)^\frac{1}{q} 
\leq \left( \mathbb{E}_{(M^{(n)}, \Theta)} \left[ \left( \sum_{j=1}^n l(M_j, \Theta) \right)^2 \right] \right)^\frac{1}{2} \tag{171}
\leq \left( \mathbb{E}_{(M^{(n)}, \Theta)} \left[ \left( \sum_{j=1}^n l(M_j, \Theta) \right)^2 \right] \right)^\frac{1}{2} + \frac{d}{\sqrt{\pi}} \sum_{i=1}^d \mathbb{E}_\Theta \left[ \left( \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2 \right]^{\frac{1}{2}} + \frac{d}{\sqrt{\pi}} \sum_{i=1}^d \mathbb{E}_\Theta \left[ \mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu)) \right] \tag{169}
\leq \left( \mathbb{E}_{(M^{(n)}, \Theta)} \left[ \left( \sum_{j=1}^n l(M_j, \Theta) \right)^2 \right] + d \mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu)) \right)^\frac{1}{2} \tag{169}
$$

Substituting (171) in (166), produces the desired lower bound

$$
\sup_{\Theta \in \Theta} \mathbb{E}_{M^{(n)}|\Theta} \left[ W^p_{\Phi}(f(x|\hat{\Theta}(M^{(n)})), f(x|\Theta)) \mid \Theta \right] \geq \sum_{j=1}^d \left( \sum_{i=1}^d \mathbb{E}_\Theta \left[ \frac{\partial}{\partial \theta_i} \mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu)) \right] \right) \leq \sum_{j=1}^d \left( \sum_{i=1}^d \mathbb{E}_\Theta \left[ \left( \sum_{j=1}^n l(M_j, \Theta) \right)^2 \right] \right)^\frac{1}{2} \sum_{i=1}^d \mathbb{E}_\Theta \left[ \left( \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2 \right]^{\frac{1}{2}} + \sum_{j=1}^d \mathbb{E}_\Theta \left[ \mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu)) \right] \tag{175}
$$

For $l(M_j, \Theta)$ with $\mathbb{E}_{M_j|\Theta}[l(M_j, \Theta)] = 0$ and independent, by the Marcinkiewicz-Zygmund inequality in the form of (2) of [18], there exists a constant $B_2 = 1 [19]$, such that

$$
\mathbb{E}_{M^{(n)}|\Theta} \left[ \left( \sum_{j=1}^n l(M_j, \Theta) \right)^2 \mid \Theta \right] \leq B_2 \mathbb{E}_{M^{(n)}|\Theta} \left[ \sum_{j=1}^n l^2(M_j, \Theta) \mid \Theta \right] = B_2 \sum_{j=1}^n \mathbb{E}_{M^{(n)}|\Theta} \left[ \left( \sum_{i=1}^d \frac{\partial}{\partial \theta_i} \log \mu(M_j | \Theta) \right)^2 \right] \tag{170}
$$

where, by the independence of $M_j$, the expectation of each element of the summation is identical and this leads to the term $n$ in (a), which also follows from the definition of $l(M_j, \Theta)$; (b) is given by the inequality $\sum_{i=1}^n x_i^2 \leq d \sum_{i=1}^n x_i^2$; $x_i > 0$, required in order to pass the summation inside the expectation and obtain the trace of the Fisher information matrix for $M_j$.

Substituting (170) in (169), we obtain

$$
\left( \mathbb{E}_\Theta \left[ \left( \mathbb{E}_{M^{(n)}|\Theta} \left[ \left| g(M^{(n)}, \Theta) \right|^2 \mid \Theta \right] \right)^\frac{q}{2} \right] \right)^\frac{1}{q} \leq \left( \mathbb{E}_{(M^{(n)}, \Theta)} \left[ \left( \sum_{j=1}^n l(M_j, \Theta) \right)^2 \right] \right)^\frac{1}{2} + \frac{d}{\sqrt{\pi}} \sum_{i=1}^d \mathbb{E}_\Theta \left[ \left( \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2 \right]^{\frac{1}{2}} + \frac{d}{\sqrt{\pi}} \sum_{i=1}^d \mathbb{E}_\Theta \left[ \mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu)) \right] \tag{171}
$$

H. Proof of Corollary 3

Using the expression of the score function,

$$
S_\Theta(X) = \frac{\partial}{\partial \theta} \log f(x|\Theta) = \frac{1}{2\sigma^2} (x - \theta),
$$

we can compute

$$
\Omega_X^{(p)}(\theta) = \frac{1}{\sigma^2} \sum_{i=1}^d \left( \mathbb{E}_{X|\Theta} \left[ \left| S_\Theta(X) \right|^\frac{p}{2} \mid \Theta = \theta \right] \right)^{p-1} \tag{171}
= \left( \frac{1}{\sigma^2} \right)^p \sum_{i=1}^d \left( \mathbb{E}_{X|\Theta} \left[ \left| X_i - \theta_i \right|^\frac{p}{2} \mid \Theta = \theta \right] \right)^{p-1} \tag{171}
= \left( \frac{1}{\sigma^2} \right)^p \left( 1 + \frac{1}{\sqrt{2\pi\sigma^2}} \right)^{-1} \times \sum_{i=1}^d \left( \int _{x_i} \left| x_i - \theta_i \right|^\frac{p}{2} \exp \left[ -\frac{(x_i - \theta_i)^2}{2\sigma^2} \right] dx_i \right)^{-1} \tag{171}
$$

Using the gamma function, the above integral becomes

$$
\int _{x_i} \left| x_i - \theta_i \right|^\frac{p}{2} \exp \left[ -\frac{(x_i - \theta_i)^2}{2\sigma^2} \right] dx_i = \frac{2 \pi \sigma^2}{\Gamma \left( \frac{p+1}{2} \right)} \frac{1}{\sqrt{2\pi\sigma^2}} \tag{171}
$$

Then, we obtain further

$$
\Omega_X^{(p)}(\theta) = \frac{d}{\sqrt{\pi}} \sigma \left[ \frac{\Gamma \left( \frac{p+1}{2} \right)}{\Gamma \left( \frac{p+1}{2} \right)} \right] \tag{171}
$$

1) Case $p \geq 2$: From Theorem 3 if $p \geq 2$, we know that

$$
\sup_{\Theta \in \Theta} \mathbb{E}_{M^{(n)}|\Theta} \left[ W^p_{\Phi}(f(x|\hat{\Theta}(M^{(n)})), f(x|\Theta)) \mid \Theta \right] \geq \sum_{j=1}^d \left( \sum_{i=1}^d \mathbb{E}_\Theta \left[ \left( \frac{\partial}{\partial \theta_i} \mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu)) \right] \left( \sum_{i=1}^d \mathbb{E}_\Theta \left[ \left( \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2 \right]^{\frac{1}{2}} + \sum_{i=1}^d \mathbb{E}_\Theta \left[ \mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu)) \right] \tag{175}
$$

For the Gaussian location model,

$$
\mathbb{E}_\Theta \left[ \left( \frac{\partial}{\partial \theta_i} \mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu)) \right] \left( \sum_{i=1}^d \mathbb{E}_\Theta \left[ \left( \frac{\partial}{\partial \theta_i} \log \mu(\Theta) \right)^2 \right]^{\frac{1}{2}} + \sum_{i=1}^d \mathbb{E}_\Theta \left[ \mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu)) \right] \tag{175}
$$

We need to compute an upper bound on $\mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu))$. If $p \geq 2$, then Theorem 3 gives us that, for some $r \geq 1$, the upper bound holds

$$
\mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu)) \leq \min \{ \mathbb{I}_{\left( 0 \right)}(\mathbb{I}(\mu)), A_{0_0} k^{-\frac{1}{2}} \} \tag{177}
$$
We move on to compute the value of $I_0$. That is, using the same approach as in Corollary 1 of [2], for $r = 2 \geq 1$, we obtain $I_0 = \frac{8}{32^2}$. Then, combining (174), (176), (177) and by Remark 3 we get

$$
sup_{\theta \in \Theta} E_{M(n)}[\Theta] \left[ \frac{d}{d\theta} \left( f(\theta) \right) \right] \geq \frac{d}{32^2} \times
$$

(178)

2) Case $1 < p < 2$: By Theorem 6 if $1 < p < 2$, we have that

$$
\sup_{\theta \in \Theta} E_{M(n)}[\Theta] \left[ \frac{d}{d\theta} \left( f(\theta) \right) \right] \geq
$$

(179)

For the Gaussian location model,

$$
E_{\Theta} \left[ \frac{d}{d\theta} \left( f(\theta) \right) \right] = \frac{1}{\theta_1} = 1.
$$

(180)

We need to compute an upper bound on $\Omega_M^{(p)}(\Theta)$. If $1 < p < 2$, then, Theorem 6 gives us that, for some $r \geq \frac{1}{p-1}$, the upper bound holds

$$
\Omega_M^{(p)}(\Theta) \leq \min \left\{ \left( \gamma_X(\theta), d \frac{2^2}{\pi^2} \right)^{\frac{1}{1-p}} \left( \frac{2}{p} \right)^{\frac{2}{p} k} \right\}.
$$

(181)

Substituting (180) in (179) and by Remark 3, for $p > 1.5$, which is required for the Beta function $B(\cdot)$ to exist, we obtain

$$
I^{(p)}(\mu_i) = \frac{\pi}{2} \left( \frac{2}{B} \right)^{p-1} \left[ B \left( 2p-1,2p-3 \right) \right]^{\frac{1}{p-1}}
$$

and

$$
sup_{\theta \in \Theta} E_{M(n)}[\Theta] \left[ \frac{d}{d\theta} \left( f(\theta) \right) \right] \geq \frac{d}{32^2} \times
$$

$$
\left( \sqrt{n} d \frac{1}{\sigma} (p-1) \right)^{\frac{1}{p}} \left[ \left( \frac{\pi^2}{p\sigma^2} \right) \right]^{\frac{p}{p-1}} +
$$

and

$$
\left( \sqrt{n} (p-1) \right)^{\frac{p+k(2-p)}{p}} \left[ \left( \frac{\pi^2}{p\sigma^2} \right) \right]^{\frac{2}{p} k} +
$$

$$
\left( \sqrt{n} (p-1) \right)^{\frac{p}{p-1}} \left( \frac{8k}{3\sigma^2} \right)^{\frac{1}{p}} \right).
$$

A. Auxiliary results

Lemma 1 (Extension of Lemma 1 of [2] to $p$-norms). For $p \geq 1$ and the parameter $\theta = [\theta_1, \ldots, \theta_d]$, $i = 1 : d$, the $(i, i)$th element of the generalized Fisher information matrix is lower bounded by

$$
[I_\theta^{(p)}(\theta)]_{ii} \geq \left( E_{M(\theta)} \left[ \left( \frac{\theta}{[S_\theta(X)|M]} \right)^{\frac{p}{p-1}} \right] \right)^{\frac{p}{p-1}}.
$$

Proof: The $(i, i)$th element of the generalized Fisher information matrix of order $p \geq 1$ associated to $M$ is equal to

$$
[I_\theta^{(p)}(\theta)]_{ii} = \left( E_{M(\theta)} \left[ |S_\theta(X)|M \right] \right)^{\frac{p}{p-1}}.
$$

We start lower bounding the score function as

$$
S_\theta(m) = \left( \frac{1}{p(m)(\theta)} \right)^{\frac{1}{p-1}} \frac{\partial}{\partial \theta_i} \left( \int x f(x) \right) \frac{1}{p(m)(\theta)} \partial \theta_i \left( \log f(x) \right) dx.
$$

Taking the absolute value, raising both sides to the power $\frac{p}{p-1}$, taking the expectation with respect to $M(\theta)$ and raising again everything to the power $p - 1$, we obtain the desired result

$$
[I_\theta^{(p)}(\theta)]_{ii} = \left( E_{M(\theta)} \left[ \left( \frac{\theta}{[S_\theta(X)|M]} \right)^{\frac{p}{p-1}} \right] \right)^{\frac{p}{p-1}}.
$$

Lemma 2 (Extension of Lemma 2 of [2] to $p$-norms, $1 \leq p < 2$). For $1 \leq p < 2$, $\theta = [\theta_1, \ldots, \theta_d]$, the trace of the generalized Fisher information matrix is lower bounded by

$$
\Omega_\Theta^{(p)}(\theta) \leq \sum_{j=1}^{d} p^{p-1}(m_j) \left| \left| E(X|\theta, m_j) \right| \right|_{p}.
$$

Proof: We begin with the definition of the trace of a matrix and we apply Lemma 1 yielding

$$
\Omega_\Theta^{(p)}(\theta) = \sum_{i=1}^{d} \left( I_\theta^{(p)}(\theta) \right)_{ii}
$$

$$
= \sum_{i=1}^{d} \left( E_{M(\theta)} \left[ \left( \frac{\theta}{[S_\theta(X)|M]} \right)^{\frac{p}{p-1}} \right] \right)^{\frac{p}{p-1}}
$$

$$
= \sum_{i=1}^{d} \left( \sum_{j=1}^{d} p^{p-1}(m_j) \left| \left| E(X|\theta, m_j) \right| \right|_{p} \right)^{\frac{p}{p-1}}
$$

$$
\leq \sum_{i=1}^{d} \sum_{j=1}^{d} p^{p-1}(m_j) \left| \left| E(X|\theta, m_j) \right| \right|_{p}
$$

(182)
by a measurable function given by (14), we have that

\[ \sum_{j=1}^{2^k} p^{p-1}(m_j|\theta) \sum_{i=1}^{d} \left[ \mathbb{E}(X_i|\theta,M) \left[ S_{\theta}(X)|M \right] \right]^p \]

\[ = \sum_{j=1}^{2^k} p^{p-1}(m_j|\theta) \left[ \mathbb{E}(X_i|\theta,m_j) \left[ S_{\theta}(X)|m_j \right] \right]^p, \quad (183) \]

where (183) follows from the inequality \((\sum_{j=1}^{2^k} x_j)^{p-1} \leq \sum_{j=1}^{2^k} x_j^{p-1}\), for \(x_j > 0\) and \(p-1 < 1\), and (183) from the definition of the \(p\)-norm, \(||x||_p = \sum_{i=1}^{d} |x_i|^p\). Then, we obtain the final upper bound

\[ \Omega_M^{(p)}(\theta) \leq \sum_{j=1}^{2^k} p^{p-1}(m_j|\theta) \left[ \mathbb{E}(X_i|\theta,m_j) \left[ S_{\theta}(X)|m_j \right] \right]^p. \]

\[ \text{Theorem 5 (Extension of Theorem 2 of [2] to } p\text{-norms, } 1 \leq p < 2). \]

If for any \(\theta \in \Theta\) and any unit vector \(u \in \mathbb{R}^d\),

\[ ||\langle u, S_{\theta}(X) \rangle||_{p,\phi_x}^2 \leq I_0 \]

holds for some \(r \geq \frac{p}{2(p-1)}\), then

\[ \Omega_M^{(p)}(\theta) \leq \min \{ \Omega_X^{(p)}(\theta), d^{\frac{2-p}{p}} I_0^\frac{2}{p} (2^k)^{2-p} p^k k^{\frac{p}{2}} \}. \]

Proof: By the inequality between the generalized Fisher information associated to a random vector and that of its transformation by a measurable function given by (14), we have that

\[ \Omega_M^{(p)}(\theta) \leq \Omega_X^{(p)}(\theta). \]

From the proof of Theorem 2 of [2], with the notation \(t = p(m|\theta)\), we have that

\[ ||\mathbb{E}(X_i|\theta,m) [S_{\theta}(X)|m]||_2 \leq I_0^\frac{1}{2} \left( \log \frac{2}{t} \right)^{\frac{1}{p}}, \]

and, together with the following inequality between different norms, for any vector \(x \in \mathbb{R}^d, 0 < p < 2\),

\[ ||x||_2 \leq ||x||_p \leq d^{\frac{1}{p}-\frac{1}{2}} ||x||_2, \]

yield the upper bound on the \(p\)-norm as

\[ ||\mathbb{E}(X_i|\theta,m) [S_{\theta}(X)|m]||_p \leq d^{\frac{2-p}{p}} I_0^\frac{2}{p} \left( \log \frac{2}{t} \right)^{\frac{p}{2}}. \]

(186)

(187)

Let \(t_j = p(m_j|\theta)\). Then, Lemma 2 upper bounds the trace of the generalized Fisher information matrix

\[ \Omega_M^{(p)}(\theta) \leq \sum_{j=1}^{2^k} t_j^{p-1} \left[ \mathbb{E}(X_i|\theta,m_j) \left[ S_{\theta}(X)|m_j \right] \right]^p \]

\[ \leq d^{\frac{2-p}{p}} I_0^\frac{2}{p} \sum_{j=1}^{2^k} t_j^{p-1} \left( \log \frac{2}{t_j} \right)^{\frac{p}{2}}, \]

(189)

where the last step follows from (188). Using the same argument as in [2], let \(\phi(x)\) be the concave envelope of \(f : (0,1] \rightarrow \mathbb{R}\),

\[ f(x) = x^{p-1} \left( \log \frac{2}{x} \right)^{\frac{p}{2}}. \]

Then, by this definition and the concavity of \(\phi(x)\),

\[ \Omega_M^{(p)}(\theta) \leq d^{\frac{2-p}{p}} I_0^\frac{2}{p} \sum_{j=1}^{2^k} \phi(t_j) \leq d^{\frac{2-p}{p}} I_0^\frac{2}{p} 2^k \phi \left( \sum_{j=1}^{2^k} t_j^{-1} \right)^{\frac{p}{2}} \]

\[ = d^{\frac{2-p}{p}} I_0^\frac{2}{p} 2^k \phi \left( \frac{1}{2^k} \right) \leq d^{\frac{2-p}{p}} I_0^\frac{2}{p} \phi \left( (2^k)^{2-p} (k+1)^2 \right) \]

\[ \leq d^{\frac{2-p}{p}} I_0^\frac{2}{p} \phi \left( (k+1)^2 p^k k^{\frac{p}{2}} \right), \]

where we selected \(\phi(x) = x^{p-1} \left( \log \frac{2}{x} \right)^{\frac{p}{2}}\), which, for \(r \geq \frac{1}{p-1}\) and \(1 < p < 2\), is concave on \(x \in (0,\frac{1}{2}]\). This is true because the function \(f(x) = x \left( \log \frac{2}{x} \right)^{\frac{p}{2}}\), for \(r \geq \frac{p}{2(p-1)}\) and \(1 < p < 2\), is concave on \(x \in (0,\frac{1}{2}]\) and \(\phi(x) = f^{p-1}(x)\) and \(x \rightarrow x^{p-1}\) is non-decreasing and concave for \(p-1 < 1\) [page 84 of [22]]. For any \(r \geq 1\) and \(k \geq 1, (k+1)^2 \leq 2^k p^k\).

Lemma 3 (Extension of Lemma 2 of [2] to \(p\)-norms, \(p \geq 2\)). For \(p \geq 2\), \(\theta = [\theta_1, \ldots, \theta_d]\), the trace of the generalized Fisher information matrix is lower bounded by

\[ \Omega_M^{(p)}(\theta) \geq \sum_{j=1}^{2^k} p(m_j|\theta) \left[ \mathbb{E}(X_i|\theta,m_j) \left[ S_{\theta}(X)|m_j \right] \right]^p. \]

Proof: We begin with the definition of the trace of a matrix and apply Lemma 1, which yields

\[ \Omega_M^{(p)}(\theta) = \sum_{i=1}^{d} \left[ \mathbb{E}(X_i|\theta,M) \left[ S_{\theta}(X)|M \right] \right]^{\frac{p}{p-1}} \]

\[ \leq \sum_{i=1}^{d} \mathbb{E}(X_i|\theta,M) \left[ S_{\theta}(X)|M \right]^{\frac{p}{p-1}} \]

\[ = \sum_{i=1}^{d} \mathbb{E}(X_i|\theta,M) \left[ S_{\theta}(X)|M \right] \]

\[ = \mathbb{E}(X|M) \left[ S_{\theta}(X)|M \right], \]

where the inequality (190) is given by Jensen’s inequality for convex functions of expectations, \(\phi(x) = x^{p-1}, p > 2\), \(\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))\), and the last step (191) follows from the definition of the \(p\)-norm, \(||x||_p = \sum_{i=1}^{d} |x_i|^p\). Writing explicitly the outer expectation from (191), we obtain the final result

\[ \Omega_M^{(p)}(\theta) \geq \sum_{j=1}^{2^k} p(m_j|\theta) \left[ \mathbb{E}(X_i|\theta,m_j) \left[ S_{\theta}(X)|m_j \right] \right]^p. \]

\[ \Omega_M^{(p)}(\theta) \leq \min \left\{ \Omega_X^{(p)}(\theta), I_0^\frac{2}{p} p^k k^{\frac{p}{2}} \right\}. \]

Theorem 6 (Extension of Theorem 2 of [2] to \(p\)-norms, \(p \geq 2\)). If for any \(\theta \in \Theta\) and any unit vector \(u \in \mathbb{R}^d\),

\[ ||\langle u, S_{\theta}(X) \rangle||_{p,\phi_x}^2 \leq I_0 \]

holds for some \(r \geq \frac{p}{2}\), then

\[ \Omega_M^{(p)}(\theta) \leq \min \{ \Omega_X^{(p)}(\theta), I_0^\frac{2}{p} p^k k^{\frac{p}{2}} \}. \]
Proof: By the inequality between the generalized Fisher information associated to a random vector and that of its transformation by a measurable function given by [14], we have that
\[ \Omega_M^{(p)}(\theta) \leq \Omega_M^{(r)}(\theta). \] (194)

From the proof of Theorem 2 of [2], with the notation \( t = p(m, \theta) \), we have that
\[ \left\| E_{(\theta, m)} [S_{\theta}(X)|m] \right\|_p \leq 2 \left( \log \frac{2}{t} \right)^{1/2}, \] (195)
and, together with the following inequality between different norms, for any vector \( x \in \mathbb{R}^d \), \( p \geq 2 \),
\[ ||x||_p \leq ||x||_2 \leq d^{1/2} \frac{1}{p} ||x||_p, \] (196)
yield the upper bound on the \( p \)-norm as
\[ \left\| E_{(\theta, m)} [S_{\theta}(X)|m] \right\|_p \leq 2 \left( \log \frac{2}{t} \right)^{\frac{p}{2}}. \] (197)

Lemma 4 upper bounds the trace of the generalized Fisher matrix
\[ \Omega_M^{(p)}(\theta) \leq \sum_{j=1}^{2^k} t_j \left\| E_{(\theta, m, j)} [S_{\theta}(X)|m, j] \right\|_p^p \leq \sum_{j=1}^{2^k} t_j \left( \log \frac{2}{t_j} \right)^{\frac{p}{2}}, \] (198)
where the last step follows from (197). Using the same argument as in [2], let \( \phi(x) \) be the concave envelope of \( f : (0, 1] \rightarrow \mathbb{R} \), \( f(x) = x (\log \frac{2}{x})^r \). Then, by this definition and the concavity of \( \phi \),
\[ t_M^{(p)}(\theta) \leq \sum_{j=1}^{2^k} \phi(t_j) \leq \left( \log \frac{2}{t_j} \right)^{\frac{p}{2}} \frac{\sum_{j=1}^{2^k} t_j}{2^k} \]
\[ = \frac{\phi(\frac{1}{2^k})}{2^k} \leq \frac{\phi(k+1)}{2^k p^k}, \]
where we selected \( \phi(x) = x (\log \frac{2}{x})^r \), which, for \( r \geq \frac{p}{2} \) and \( p \geq 2 \), is concave on \( x \in (0, \frac{1}{2}] \).

REFERENCES

[1] Y. Han, A. Özgür and T. Weissman, "Geometric lower bounds for distributed parameter estimation under communication constraints," Proceedings of the 31st Conference On Learning Theory (COLT), 75:3163-3188, 2018.
[2] L. P. Barnes, Y. Han and A. Özgür, "Lower bounds for learning distributions under communication constraints via Fisher information," arXiv:1902.02890, 2019.
[3] S. Kamath, A. Orlitovsky and V. Pichapati and A.-T. Suresh, "On learning distributions from their samples," JMLR Workshop and Conference Proceedings, Vol. 40, pp. 1-35, 2015.
[4] Y. Han, P. Mukherjee, A. Özgür and T. Weissman, "Distributed statistical estimation of high-dimensional and nonparametric distributions," Proceedings of the IEEE International Symposium on Information Theory (ISIT), 2018.
[5] S.-I. Amari, "On optimal data compression in multiterminal statistical inference," IEEE Transactions on Information Theory, Vol. 57, N0. 9, pp. 5577-5587, 2011.
[6] C. Villani, "Optimal transport old and new," Springer-Verlag Berlin Heidelberg, 2009.
[7] H.-L. van Trees, "Detection, Estimation and Modulation Theory. Part I," in Wiley & Sons, 1968.
[8] S. Sarbu and A. Zaidi, "On learning parametric distributions from quantized samples" Draft. Available at [http://www-idsyscom.univ-mlv.fr/~zaid/publications/proofs-paper-isit2021.pdf], 2021.
[9] C. Froger, C. Zhang, H. Mobahi, M. Araya-Polo, T. Poggio, "Learning with a Wasserstein loss," in Advances in Neural Information Processing Systems 28, 2015.
[10] L. Ambrogioni, U. Güçlü, Y. Güçlüturk, M. Hinne, E. Maris, M.A.J. van Gerven, "Wasserstein variational inference," Advances in Neural Information Processing Systems 31, 2018.
[11] I. Tolstikhin, O. Bousquet, S. Gelly, B. Schölkopf, "Wasserstein autoencoders," Proceedings of the 6th International Conference on Learning Representations, 2018.
[12] M. Arjovsky, S. Chintala, L. Bottou, "Wasserstein generative adversarial networks," Proceedings of the 34th International Conference on Machine Learning, 2017.
[13] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, "Improved training of Wasserstein GANs," Advances in Neural Information Processing Systems 30, 2017.
[14] D.E. Boekee, "Generalized Fisher information with application to estimation problems," in IFAC Workshop on Information and Systems, 10:75-82, 1977.
[15] D.E. Boekee, "A generalization of the Fisher information measure," Ph.D. Thesis, Dept. of El. Eng., Delft Univ. of Tech., Delft, The Netherlands, 1977.
[16] R.D. Gill and B.Y. Levit, "Applications of the van Trees inequality: a Bayesian Cramér-Rao bound," in Bernoulli, 1:59-79, 1995.
[17] B. Yu, "Assouad, Fano, and Le Cam," Chapter in D. Pollard, E. Torgersen, G.L. Yang(eds), Festschrift for Lucien Le Cam, Springer, New York, NY, 1997.
[18] Y.-F. Ren and H.-Y. Liang, "On the best constant in Marcinikiewicz-Zygmund inequality," in Statistics & Probability Letters, 53, 227-233, 2001.
[19] D.L. Burkholder, "Sharp inequalities for martingales and stochastic integrals", in Astérisque, tome 157-158, p. 75-94, 1988.
[20] V. V. Gyory, "Introduction to the non-asymptotic analysis of random matrices", in arXiv preprint. arxiv:1011.3027, 2010.
[21] Z. Goldfeld, K. Greenewald, J. Niels-Weed and Y. Polyanskiy, "Convergence of smoothed empirical measures with applications to entropy estimation," in IEEE Transactions on Information Theory, vol. 66, 7:4368-4391, July 2020.
[22] S. Boyd and L. Vandenberghe, "Convex Optimization," Cambridge University Press, 2004.