2023 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-7554-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISIT54713.2023.10206980

L^1 Estimation in Gaussian Noise: On the Optimality of Linear Estimators

Leighton P. Barnes*, Alex Dytso†, and H. Vincent Poor‡
*Center for Communications Research, Princeton, NJ 08540, USA, leightonbarnes@gmail.com
†Qualcomm Flarion Technology, Inc., Bridgewater, NJ 08807, USA, odytso2@gmail.com
‡Princeton University, Princeton, NJ 08544, USA, poor@princeton.edu

***Abstract*—Consider the problem of estimating a random variable X in Gaussian noise under L^1 fidelity criteria. It is well-known that in the L^1 setting, the optimal Bayesian estimator is given by the conditional median. The goal of this work is to characterize the set of prior distributions on X for which the conditional median corresponds to a linear estimator. This work shows that neither discrete nor compactly supported distributions can induce a linear conditional median. Moreover, under certain non-trivial restrictions on the set of allowed probability distributions, the Gaussian is shown to be the only solution that induces a linear conditional median.**

I. Introduction

Consider the problem of estimating a scalar random variable X from a noisy observation

$$Y = X + Z \tag{1}$$

where Z is *standard normal* independent of X. In this work, we are interested in studying the *condition median estimator* of X given noisy observation Y, which is the optimal Bayesian estimator under the L^1 error criterion.

To properly define the conditional median, recall that for a random variable U the *quantile function* or the *inverse cumulative distribution function* (cdf) is defined as

$$F_U^{-1}(p) = \inf\{x \in \mathbb{R} : p \leq F_U(x)\}, \, p \in (0, 1), \tag{2}$$

where F_U is the cdf of U. The conditional median is then defined as

$$\mathsf{m}(X|Y = y) = F_{X|Y=y}^{-1}\left(\frac{1}{2}\right), \, y \in \mathbb{R}, \tag{3}$$

where $F_{X|Y=y}$ is the conditional cdf of X given $Y = y$.

In estimation theory, the conditional median is of interest since it is the optimal estimation of X under *the mean absolute deviation error* criterion (a.k.a. L^1 estimation error):

$$\mathsf{m}(X|Y) \in \arg \min_{f: \, \mathbb{E}[|f(Y)|] < \infty} \mathbb{E}[|X - f(Y)|]. \tag{4}$$

The condition median maps the observation $Y \in \mathbb{R}$ to the support of X; that is[1]

$$\mathsf{m}(X|Y) : \mathbb{R} \to \mathsf{supp}(X). \tag{5}$$

This is an appealing property of the conditional median since, in contrast, *the conditional mean*, which is the optimal estimator under the L^2 error criterion and is defined as

$$\mathbb{E}[X|Y = y] = \int x \, \mathrm{d}P_{X|Y=y}(x), \, y \in \mathbb{R}, \tag{6}$$

maps the observation Y to the interior of the convex hull of $\mathsf{supp}(X)$.

A. The Problem Setup

In estimation theory, *linear estimators* play an important role. In particular, it is important to understand under which criteria the optimal Bayesian estimators are linear functions of the observation Y. Despite considerable research into linear estimators, to the best of our knowledge, the question of identifying the set of prior distributions on X that ensure that $\mathsf{m}(X|Y)$ is a linear function of Y has not been characterized. In this work, we seek to close this gap.

Formally, we seek to answer the following question: *Which distributions on the input X ensure that there exists some constant a such that for all $y \in \mathbb{R}$*

$$\mathsf{m}(X|Y = y) = ay? \tag{7}$$

B. Prior Work

Conditions for the optimality of linear estimators have received considerable attention for squared error loss under which the optimal estimator is given by the conditional mean [1]. In particular, for the case when $P_{Y|X}$ belongs to an *exponential family*, it is well known that the conditional mean is linear if and only if X is distributed according to a conjugate prior [2], [3] in which case $a = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}$. Moreover, for the Gaussian and Poisson noise models, we even have stability results that show, if the conditional expectation is close to a linear function in the L^2 distance, then the distribution of X needs to be close to a matching prior (Gaussian for Gaussian noise and gamma for Poisson noise) in the Lévy distance [4], [5].

For additive noise channels, i.e., $Y = X + N$ where N is not necessarily Gaussian, the authors of [6] characterized necessary and sufficient conditions for the linearity of the optimal Bayesian estimators for the case of L^p Bayesian risks (i.e., $\mathbb{E}[|X - f(Y)|^p]$) with p taking only *even* values. More specifically, the authors of [6] found the characteristic function of X as a function of the characteristic function of N. These

[1] Let P_X be the probability distribution of X. Then, we say $x \in \mathsf{supp}(X)$ if and only if $P_X(\mathcal{S}) > 0$ for every $\mathcal{S}$ containing x.

results, however, do not generalize to our case of $p = 1$, which, as will be shown, is considerably more difficult.

The conditional median plays an important role in our analysis, and for a detailed study of the properties of the conditional median in the most abstract σ-algebra setting, the interested reader is referred to [7], [8]. For recent applications of the conditional median, the interested reader is referred to [9] and references therein.

C. Outline and Contribtuions

The paper outline and contributions are as follows:

1) In Section II, we make some preliminary observations about the solution to (7) and show:
   - Section II-A, Proposistion 1, shows that a Gaussian distribution satisfies (7);
   - Section II-B, Proposition 2 provides an equivalent condition to (7) in terms of an integral equation;
   - Section II-C shows that finding a solution to (7) is equivalent to characterizing the null space of a certain integral operator; and
   - Section II-D seeks to point out some of the difficulties with finding the uniqueness of the solution to (7).
2) In Section III, we present our main results and show:
   - Section III-A, Proposition 4, shows that the distribution that satisfies (7) needs to be fully supported;
   - Section III-B, Theorem 5 shows that the only admissible values of a lie in $[0, 1)$ and Proposition 6 provides a characterization of a in terms of the pdf of X; and
   - Section III-C, Theorem 7, shows that if we restrict our attention to only two types of Gaussian mixtures (i.e., finite Gaussian mixtures and arbitrary Gaussian mixtures with a fixed variance), then Gaussian is the only solution to (7).
3) Section IV concludes the paper.

Some of the proofs a relegated to the extended version of the paper [10].

Notation: The pdf of a zero mean Gaussian distribution with variance σ^2 is denote by ϕ_{σ^2}. We also let $\phi = \phi_1$ (i.e., the pdf of standard normal).

II. Preliminary Observations

In this section, we begin by providing preliminary observations about the problem, derive a necessary and sufficient condition for (7) to hold, and try to point out the reason why solving the problem is a challenging task. Along the way, we also derive some results that might be of independent interest.

A. Gaussian X is a Solution

We begin by showing that the set of distributions that satisfies (7) is not empty.

Proposition 1. *If $0 \le a < 1$, a Gaussian random variable $X \sim \mathcal{N}(0, \sigma_X^2)$ satisfies (7) if $\sigma_X^2 = \frac{a}{1-a}$.*

Proof: Suppose that $X \sim \mathcal{N}(0, \sigma_X^2)$; then the conditional distribution $X|Y = y \sim \mathcal{N}(\frac{\sigma_X^2}{1+\sigma_X^2} y, \frac{\sigma_X^2}{1+\sigma_X^2})$. Since Gaussian distributions are symmetric, the conditional median and conditional mean coincide, and we have that

$$\mathsf{m}(X|Y = y) = \mathbb{E}[X|Y = y] = \frac{\sigma_X^2}{1 + \sigma_X^2} y. \tag{8}$$

Solving for σ_X^2 concludes the proof. ■

In Proposition 1, for the case of $a = 0$, and for the rest of the paper, we do not distinguish between point measures and Gaussian measures with zero variance and treat them as the same objects.

Remark 1. *We conjecture that the Gaussian distribution the zero mean and $\sigma_X^2 = \frac{a}{1-a}$ is the unique solution to (7). The supporting arguments for this conjecture will be given in Section III.*

B. An Equivalent Condition

In this subsection, we derive a condition that is equivalent to (7). Our starting place is the following condition akin to the orthogonality principle [6], [11]: a function f minimizes (4) if and only if

$$\mathbb{E}[\mathsf{sign}(X - f(Y))g(Y)] = 0, \tag{9}$$

for all g such that $\mathbb{E}[|g(Y)|] < \infty$.

Proposition 2. *X satisfies (7) if and only if for all $y \in \mathbb{R}$*

$$\mathbb{E}[\mathsf{sign}(X - ay)\phi(y - X)] = 0. \tag{10}$$

Proof: We seek to show that for $f(Y) = aY$, the condition in (9) is equivalent to (10). Note that (10) can be equivalently re-written as: for all g such that $\mathbb{E}[|g(Y)|] < \infty$

$$0 = \mathbb{E}[\mathsf{sign}(X - aY)g(Y)] \tag{11}$$
$$= \mathbb{E}[\mathbb{E}[\mathsf{sign}(X - aY)|Y]g(Y)] \tag{12}$$
$$= \mathbb{E}[h(Y)g(Y)], \tag{13}$$

where we have defined $h(Y) = \mathbb{E}[\mathsf{sign}(X - aY)|Y]$. The fact that (13) is equivalent to

$$0 = h(y) \text{ for all } y \in \mathbb{R}, \tag{14}$$

is a standard fact (see, for example, [12, Lem. 10.1.1]). This concludes the proof. ■

C. Operator Theory Perspective

Consider the following integral operator on the set of L^1 functions:

$$T_a[f](y) = \int_{-\infty}^{\infty} K_a(x, y)f(x)\mathrm{d}x \tag{15}$$

where the *kernel* $K_a(x, y)$ is given by

$$K_a(x, y) = \mathsf{sign}(x - ay)\phi(y - x). \tag{16}$$

If we restrict our attention only to random variables X having pdfs, finding the set of solutions to (10) is equivalent to characterizing the null space of $T_a[f]$ over the space L^1_+; that is

$$\mathcal{N}(T_a) = \left\{ f : f \ge 0, \int f < \infty, T_a[f] = 0 \right\}. \tag{17}$$

Although we are only interested in L^1 functions (for f to be a pdf), the operator T_a can also be thought of as a bounded linear operator from $L^p(\mathbb{R})$ to itself for any $1 \le p \le \infty$.

D. What are the Challenges?

We now discuss some of the challenges and why some standard approaches no longer work.

1) Techniques for Showing Linearity of the Conditional Expectation Are Not Applicable: It is well known that for the conditional expectation

$$\mathbb{E}[X|Y = y] = ay, \forall y \in \mathbb{R}, \tag{18}$$

if and only if $a \in [0, 1)$ and $f_X = \phi_{\frac{a}{1-a}}$. The authors of this paper are aware of five district ways of showing this fact; four of these methods, some of which are new, are provided in the extended version of the paper [10]. However, none of these techniques appear to be generalizable to the conditional median setting.

2) Construction of a Solution Through Symmetry Arguments Will Not Work: The proof of Proposition 1 relied on the fact that if X is Gaussian, then $X|Y = y$ is a symmetric distribution for all y [2] and, hence, the mean and the median coincide. The next result, which might be of independent interest, shows that this construction works only in the Gaussian case.

Theorem 3. *If X is Gaussian, then $X|Y = y$ is symmetric for all y. Conversely, if $X|Y = y$ is symmetric for all $y \in S$ where S is a subset of $\mathbb{R}$ that has an accumulation, then X is Gaussian.*

III. Main Results

In this section, we present our main results. We begin by presenting some simple statements about the support of the distribution that needs to satisfy (7).

A. On the Support of the Distribution

Proposition 4. *Suppose that X satisfies* (7). *Then,*
- *if $a = 0$, then $X = 0$ a.s.*
- *if $a \neq 0$, then the* $\mathsf{supp}(X) = \mathbb{R}$ *(i.e., X is fully supported and unbounded).*

Proof: If the conditional median is a constant, then so is the underlying random variable (see, for example, [7, Thm. 3]). To show the second property recall that

$$\mathsf{m}(X|Y) : \mathbb{R} \to \mathsf{supp}(X), \tag{19}$$

However, since $\mathsf{m}(X|Y = y) = ay$, $y \in \mathbb{R}$, the range is clearly given by $\mathbb{R}$. Therefore, $\mathsf{supp}(X) = \mathbb{R}$. ■

It is important to note that Proposition 4, for example, completely eliminates discrete distributions as possible solutions or distributions with compact support. Note, however, it does not preclude the possibility of mixed distributions (i.e., distributions that have both continuous and discrete components).

[2]The random variable U is said to have symmetric distribution if there exists a constant c such that $U + c \stackrel{d}{=} -(U + c)$ where $\stackrel{d}{=}$ denotes equality in distribution.

B. On the Admissible Values of a

We next show that the admissible values of a that satisfy (7) must be in $[0, 1)$.

Theorem 5.

$$\min_{a\in\mathbb{R}} \mathbb{E}[|X - aY|] = \min_{a\in[0,1)} \mathbb{E}[|X - aY|]. \tag{20}$$

In other words, the admissible values of a lie in $[0, 1)$.

Proof: Let

$$f(a) = \mathbb{E}[|X - aY|]. \tag{21}$$

Then,

$$\begin{aligned} f'(a) &= -\mathbb{E}[\mathsf{sign}(X - aY)Y] && (22)\\ &= \mathbb{E}[\mathsf{sign}(aY - X)Y] && (23)\\ &= \mathbb{E}[\mathsf{sign}((a-1)X + aZ)(X + Z)]. && (24) \end{aligned}$$

We will show that the function $f(a)$ is *non-decreasing* for $a \ge 1$ and *non-increasing* for $a < 0$. Thus, we will reduce our search space to $a \in [0, 1]$. To aid our proof recall the FKG inequality (see for example [13]): for two non-decreasing functions f and g, we have that

$$\mathbb{E}[f(X)g(X)] \ge \mathbb{E}[f(X)]\mathbb{E}[g(X)], \tag{25}$$

or equivalently, if f is non-decreasing and g is non-increasing, then

$$\mathbb{E}[f(X)g(X)] \le \mathbb{E}[f(X)]\mathbb{E}[g(X)]. \tag{26}$$

Now, assume that $a \ge 1$; then

$$\begin{aligned} f'(a) &= \mathbb{E}[\mathsf{sign}((a-1)X + aZ)(X + Z)]\\ &= \mathbb{E}[\mathbb{E}[\mathsf{sign}((a-1)X + aZ)(X + Z)|Z]] && (27)\\ &\ge \mathbb{E}[\mathbb{E}[\mathsf{sign}((a-1)X + aZ)|Z]\mathbb{E}[X + Z|Z]] && (28)\\ &= \mathbb{E}[\mathbb{E}[\mathsf{sign}((a-1)X + aZ)|Z]Z] && (29)\\ &= \mathbb{E}[\mathsf{sign}((a-1)X + aZ)Z] && (30)\\ &= \mathbb{E}[\mathbb{E}[\mathsf{sign}((a-1)X + aZ)Z|X]] && (31)\\ &\ge \mathbb{E}[\mathbb{E}[\mathsf{sign}((a-1)X + aZ)|X]\mathbb{E}[Z|X]] && (32)\\ &= 0, && (33) \end{aligned}$$

where (28) follows by using the fact that, given Z the functions $f(X) = \mathsf{sign}((a-1)X + aZ)$ and $g(X) = X + Z$ are non-decreasing and applying the FKG inequality; (29) follows by using the fact that X and Z are independent and the assumption that $\mathbb{E}[X] = 0$ which implies that $\mathbb{E}[X + Z|Z] = Z + \mathbb{E}[X] = Z$; (32) follows by using the fact that given X the functions $f(Z) = \mathsf{sign}((a-1)X + aZ)$ and $g(Z) = Z$ are non-decreasing and applying the FKG inequality.

Now, assume that $a \le 0$; then

$$\begin{aligned} f'(a) &= \mathbb{E}[\mathsf{sign}((a-1)X + aZ)(X + Z)]\\ &= \mathbb{E}[\mathbb{E}[\mathsf{sign}((a-1)X + aZ)(X + Z)|Z]] && (34)\\ &\le \mathbb{E}[\mathbb{E}[\mathsf{sign}((a-1)X + aZ)|Z]\mathbb{E}[X + Z|Z]] && (35)\\ &= \mathbb{E}[\mathbb{E}[\mathsf{sign}((a-1)X + aZ)|Z]Z] && (36)\\ &= \mathbb{E}[\mathsf{sign}((a-1)X + aZ)Z] && (37) \end{aligned}$$

$$= \mathbb{E}[\mathbb{E}[\mathsf{sign}((a-1)X + aZ)Z|X]] \quad (38)$$
$$\le \mathbb{E}[\mathbb{E}[\mathsf{sign}((a-1)X + aZ)|X]\mathbb{E}[Z|X]] \quad (39)$$
$$= 0, \quad (40)$$

where in (35) we have used that, given Z, $f(X) = \mathsf{sign}((a-1)X + aZ)$ is non-increasing and $g(X) = X + Z$ are non-decreasing; and in (39) we have used that, given X, $f(Z) = \mathsf{sign}((a-1)X + aZ)$ is non-increasing and $g(Z) = Z$ are non-decreasing and applyed the FKG inequality.

Thus, we can assume that $a \in [0, 1]$. To conclude the proof note that by independence, we have that

$$f'(1) = \mathbb{E}[\mathsf{sign}(Z)(X + Z)] = \mathbb{E}[|Z|] > 0, \quad (41)$$

which implies that we can eliminate the value of $a = 1$. ■

Note the solution to (7) is a pair (a, f_X). The next result identifies the slope a as a function of f_X. Thus, if we have a candidate solution in terms of the pdf f_X, we no longer need to guess the value of $a \in (0, 1)$ and can determine it exactly.

Proposition 6. *Suppose that in addition to satisfying* (7) X *has a pdf* f_X*. Then, the following hold:*

- $f_X(0) > 0$*; and*
- *the value of* a *is given by*

$$a = \frac{1}{2f_X(0)\phi(0)} \int_{-\infty}^{\infty} |x|\phi(x)f_X(x)\mathrm{d}x. \quad (42)$$

Proof: If X has a pdf, then (10) can be written as: for all $y \in \mathbb{R}$

$$0 = \int_{-\infty}^{\infty} \mathsf{sign}(x - ay)\phi(y - x)f_X(x)\mathrm{d}x. \quad (43)$$

Taking the derivative of both sides of (43) we arrive at

$$0 = \frac{\mathrm{d}}{\mathrm{d}y} \int_{-\infty}^{\infty} \mathsf{sign}(x - ay)\phi(y - x)f_X(x)\mathrm{d}x \quad (44)$$
$$= -a \int_{-\infty}^{\infty} 2\delta(x - ay)\phi(y - x)f_X(x)\mathrm{d}x$$
$$+ \int_{-\infty}^{\infty} \mathsf{sign}(x - ay)(x - y)\phi(y - x)f_X(x)\mathrm{d}x \quad (45)$$
$$= -a\phi((1 - a)y)f_X(ay)$$
$$+ \int_{-\infty}^{\infty} \mathsf{sign}(x - ay)(x - y)\phi(y - x)f_X(x)\mathrm{d}x. \quad (46)$$

Setting $y = 0$ in (46), we arrive at

$$2a\phi(0)f_X(0) = \int_{-\infty}^{\infty} |x|\phi(x)f_X(x)\mathrm{d}x. \quad (47)$$

Now since the right-hand side of (47) is positive and $a \neq 0$ by Proposition 4, this implies that $f_X(0) > 0$. Now solving (47) for a concludes the proof. ■

C. Solution for Gaussian Mixtures

Let $\mathcal{P}$ be the set of all probability measures on $\mathbb{R}$. Now for a fixed $\sigma > 0$ define

$$\mathcal{P}_{\sigma^2} = \{f_X : f_X(x) = \mathbb{E}[\phi_{\sigma^2}(x - U)], U \sim P_U \in \mathcal{P}\} \quad (48)$$

In words, $\mathcal{P}_{\sigma^2}$ is the set of all Gaussian mixtures with a fixed variance σ^2. The random variable U is known as the mixing random variable. Note that we make no restrictions on the distribution P_U.

We also define the set of finite N-Gaussian mixture where both the variance and mean are allowed to vary: for a positive integer N

$$\mathcal{P}^N = \left\{ f_X : f_X(x) = \sum_{i=1}^N p_i \phi_{\sigma_i^2}(x - \mu_i), \right.$$
$$\left. p_i \ge 0, \sum_{i=1}^N p_i = 1, \sigma_i > 0, \mu_i \in \mathbb{R} \right\}. \quad (49)$$

The main result of this section is the following theorem.

Theorem 7. *Fix some* $a \in (0, 1)$*. For every finite positive integer* N*,* $f_X = \phi_{\frac{a}{1-a}}$ *is the unique solution to* (7) *over the set*

$$\mathcal{P}_{\frac{a}{1-a}} \cup \mathcal{P}^N. \quad (50)$$

Note that the set $\mathcal{P}_{\frac{a}{1-a}} \cup \mathcal{P}^N$ is a non-trivial subset of the set of all pdfs. For example, the set $\mathcal{P}^N$ for large enough N can approximate any distribution in the Lévy distance to an arbitrary degree.

Before showing the proof Theorem 7, we will provide several auxiliary results. The first result shows how Gaussian functions are mapped forward by the operator T_a.

Lemma 8. *Let* $f(x) = \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ *and let* $b = 1 + \frac{1}{\sigma^2}$*. Then,*

$$T_a[f](y) = \psi(y; a, \sigma^2, \mu), \quad (51)$$

where

$$\psi(y; a, \sigma^2, \mu)$$
$$= \mathrm{e}^{-\frac{y^2}{2} - \frac{\mu^2}{2\sigma^2}} \sqrt{\frac{1}{b}} \mathrm{e}^{\frac{(y + \frac{\mu}{\sigma^2})^2}{2b}} \mathrm{erf}\left(\frac{(1 - ba)y + \frac{\mu}{\sigma^2}}{\sqrt{2b}}\right). \quad (52)$$

Furthermore, if $a = \frac{1}{b}$ *(i.e.,* $\sigma^2 = \frac{a}{1-a}$*), then*

$$T_a[f](y) = c(a, \mu) \exp\left(-(1 - a)\frac{(y - \mu)^2}{2}\right), \quad (53)$$

where $c(a, \mu) = \sqrt{a}\mathrm{erf}\left(\frac{\mu}{\sqrt{2a\frac{a}{1-a}}}\right)$.

It is trivial to show that the operator T_a commutes with finite mixtures. The next result shows that the operator T_a also commutes with arbitrarily fixed variance mixtures.

Lemma 9. *For* $a \in (0, 1)$ *and* $0 \le \sigma^2 \le \infty$ *with* $b = 1 + \frac{1}{\sigma^2}$*. Then, for* $f_X \in \mathcal{P}_{\sigma^2}$

$$T_a[f_X](y) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathbb{E}\left[\psi(y; a, \sigma^2, U)\right]. \quad (54)$$

Moreover, for $b = \frac{1}{a}$ *we have that*

$$T_a[f_X](y) = \frac{\sqrt{a}}{\sqrt{2\pi\frac{a}{1-a}}} \mathbb{E}\left[\mathrm{e}^{-(1-a)\frac{(y-U)^2}{2}} \mathrm{erf}\left(\frac{U}{\sqrt{2\frac{1}{a}\frac{a}{1-a}}}\right)\right]. \quad (55)$$

Proof: By Fubini's theorem we have to show that $\int\int|K_a(x,y)\phi_{\sigma^2}(x-u)|\mathrm{d}P_U(u)\mathrm{d}x < \infty$. Towards that end, note that

$$\int\int|K_a(x,y)\phi_{\sigma^2}(x-u)|\mathrm{d}P_U(u)\mathrm{d}x$$
$$= \int\int \phi(y-x)\phi_{\sigma^2}(x-u)\mathrm{d}x\mathrm{d}P_U(u) \quad (56)$$
$$= \int \phi_{1+\sigma^2}(u-y)\mathrm{d}P_U(u) \le \frac{1}{\sqrt{2\pi(1+\sigma^2)}}, \quad (57)$$

where the exchange of integration order in (56) follows by Tonelli's theorem since all the arguments are non-negative. The rest of the proof follows by using Lemma 8. ■

Lemma 10. *Fix some $a \in (0,1)$ and for a finite integer N suppose that $\{(\mu_i,\sigma_i^2)\}_{i=1}^N$ are distinct with $(\mu_i,\sigma_i^2) \neq (0,\frac{a}{1-a}), \forall i$. Then, the collection of functions $\left\{\psi(y;a,\sigma_i^2,\mu_i)\right\}_{i=1}^N$ is linearly independent.*

Proof: First, note that the assumption that $(\mu_i,\sigma_i^2) \neq (0,\frac{a}{1-a})$ guarantees that $\psi(y;a,\sigma_j^2,\mu_j) \not\equiv 0$ for all i. Otherwise, the collection would be linearly dependent.

Next, we will show that the tails of $\psi(y;a,\sigma_i^2,\mu_i)$ and $\psi(y;a,\sigma_j^2,\mu_j)$ diverge, and hence we cannot write one ψ function as a linear combination of the other ψ functions. To that end, note that for $i \neq j$

$$\lim_{y\to\infty}\left|\frac{\psi(y;a,\sigma_i^2,\mu_i)}{\psi(y;a,\sigma_j^2,\mu_j)}\right|$$
$$= \lim_{y\to\infty}\left|\frac{\mathrm{e}^{-\frac{\mu_i^2}{2\sigma_i^2}}\sqrt{\frac{1}{b_i}}\mathrm{e}^{\frac{(y+\frac{\mu_i}{\sigma_i^2})^2}{2b_i}}\mathrm{erf}\left(\frac{(1-b_ia)y+\frac{\mu_i}{\sigma_i^2}}{\sqrt{2b_i}}\right)}{\mathrm{e}^{-\frac{\mu_j^2}{2\sigma_j^2}}\sqrt{\frac{1}{b_j}}\mathrm{e}^{\frac{(y+\frac{\mu_j}{\sigma_j^2})^2}{2b_j}}\mathrm{erf}\left(\frac{(1-b_ja)y+\frac{\mu_j}{\sigma_j^2}}{\sqrt{2b_j}}\right)}\right| \quad (58)$$
$$= \begin{cases} \infty & \sigma_i^2 > \sigma_j^2 \text{ or } (\sigma_i^2 = \sigma_j^2 \text{ and } \mu_i - \mu_j > 0) \\ 0 & \sigma_i^2 < \sigma_j^2 \text{ or } (\sigma_i^2 = \sigma_j^2 \text{ and } \mu_i - \mu_j < 0) \end{cases}. \quad (59)$$

This concludes the proof. ■

We are now ready to present the proof of Theorem 7.

Proof of Theorem 7: We first focus on $\mathcal{P}^N$. Suppose that $f_X \in \mathcal{P}^N$ and $T_a[f_X] = 0$; then using Lemma 10

$$0 = T_a[f_X](y) = \sum_{i=1}^N c_i\psi(y;a,\sigma_i^2,\mu_i), \quad (60)$$

where $c_i = \frac{p_i}{\sqrt{2\pi\sigma_i^2}}$. We now consider two cases. First, if $(\mu_i,\sigma_i^2) \neq (0,\frac{a}{1-a}), \forall i$, then by the linear independence of the ψ functions in Lemma 10 there exists no such f_X. Second, if there exists a j such that $(\mu_j,\sigma_j^2) = (0,\frac{a}{1-a})$, then $\psi(y;a,\sigma_i^2,\mu_i) \equiv 0$ and (60) can be written as

$$0 = \sum_{i=1:i\neq j}^N c_i\psi(y;a,\sigma_i^2,\mu_i). \quad (61)$$

By the linear independence, we have that $c_i = 0$ for all $i \neq j$. Thus, the finite mixture representing f_X must consist of a single Gaussian distribution with mean and variance given by $(\mu_j,\sigma_j^2) = (0,\frac{a}{1-a})$ and, therefore, $f_X(x) = \phi_{\frac{a}{1-a}}$.

This shows that the only solution in $\mathcal{P}^N$ is $f_X(x) = \phi_{\frac{a}{1-a}}$. We now move on to showing uniqueness over $\mathcal{P}_{\frac{a}{1-a}}$.

Let $f_X \in \mathcal{P}_{\frac{a}{1-a}}$; then by using Lemma 9, $T_a[f_X] = 0$ can be equivalently re-written as

$$0 = \mathbb{E}\left[\exp\left(-(1-a)\frac{(y-U)^2}{2}\right)\mathrm{erf}\left(\frac{U}{\sqrt{2\frac{1}{a}\frac{a}{1-a}}}\right)\right] \quad (62)$$
$$= \int \exp\left(-(1-a)\frac{(y-u)^2}{2}\right)\mathrm{d}\nu(u), \quad (63)$$

where we have defined a measure ν as

$$\mathrm{d}\nu(u) = \mathrm{erf}\left(\frac{u}{\sqrt{2\frac{1}{a}\frac{a}{1-a}}}\right)\mathrm{d}P_U(u). \quad (64)$$

Note that ν is a finite and signed measure.

The right-hand side of (63) can now be interpreted as the convolution of a Gaussian density function with the measure ν. Now taking the Fourier transform, we have that

$$0 = \exp\left(-\frac{\omega^2}{2(1-\frac{1}{b})}\right)\hat{\nu}(\omega), \quad \forall\omega \quad (65)$$

where $\hat{\nu}$ is the Fourier transform (a.k.a. characteristic function) of the measure ν; that is

$$\hat{\nu}(\omega) = \int \mathrm{e}^{i\omega u}\mathrm{d}\nu(u). \quad (66)$$

From (65) we conclude that $\hat{\nu}(\omega) = 0$, and using the fact that the Fourier transform is unique for a finite signed measure, we have that for every measurable set $\mathcal{E} \subset \mathbb{R}$

$$0 = \int_{\mathcal{E}}\mathrm{d}\nu(u) = \int_{\mathcal{E}}\mathrm{erf}\left(\frac{u}{\sqrt{2\frac{1}{a}\frac{a}{1-a}}}\right)\mathrm{d}P_U(u). \quad (67)$$

The only probability measure for which this is true is $P_U = \delta_0$. This concludes the proof.

IV. Conclusion

This work has focused on characterizing which prior distributions give an optimal estimator (with respect to L^1 loss) that is linear, which is equivalent to answering the question of when conditional medians are linear functions of the observations. We have focused on a Gaussian noise model, but the question can be considered more generally. It has been conjectured that a Gaussian prior is the only one that induces a linear conditional median, and this conjecture has been proven to hold under non-trivial restrictions on the set of allowed probability distributions. Also, along the way, several new results have been shown that might be of independent interest. For example, it has been shown that in Gaussian noise, the posterior distribution is symmetric if and only if the prior is Gaussian.

REFERENCES

[1] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer Science & Business Media, 1998.

[2] P. Diaconis and D. Ylvisaker, "Conjugate priors for exponential families," *Ann. Statist.*, pp. 269–281, 1979.

[3] J.-P. Chou, "Characterization of conjugate priors for discrete exponential families," *Statist. Sinica*, pp. 409–418, 2001.

[4] F. du Pin Calmon, Y. Polyanskiy, and Y. Wu, "Strong data processing inequalities for input constrained additive noise channels," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1879–1892, 2018.

[5] A. Dytso and H. V. Poor, "Estimation in Poisson noise: Properties of the conditional mean estimator," *IEEE Transactions on Information Theory*, vol. 66, no. 7, pp. 4304–4323, 2020.

[6] E. Akyol, K. Viswanatha, and K. Rose, "On conditions for linearity of optimal estimation," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3497–3508, 2012.

[7] R. J. Tomkins, "On conditional medians," *The Annals of Probability*, vol. 3, no. 2, pp. 375–379, 1975.

[8] Y. N. Ghosh and B. Mukherjee, "On probabilistic properties of conditional medians and quantiles," *Statistics & Probability Letters*, vol. 76, no. 16, pp. 1775–1780, 2006.

[9] D. Medarametla and E. Candès, "Distribution-free conditional median inference," *Electronic Journal of Statistics*, vol. 15, no. 2, pp. 4625–4658, 2021.

[10] L. Barnes, A. Dytso, and H. V. Poor, "L^1 estimation: On the optimality of linear estimators," 2023, draft.

[11] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, H. V. Poor, and S. S. Shitz, "On the minimum mean p th error in Gaussian noise channels and its applications," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 2012–2037, 2017.

[12] S. Resnick, *A Probability Path*. Springer, 2019.

[13] J. H. Kemperman, "On the FKG-inequality for measures on a partially ordered space," in *Indagationes Mathematicae (Proceedings)*, vol. 80, no. 4. North-Holland, 1977, pp. 313–331.

[14] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. US Government Printing Office, 1964, vol. 55.

[15] A. Dytso, H. V. Poor, and S. Shamai, "Conditional mean estimation in Gaussian noise: A meta derivative identity with applications," *IEEE Transactions on Information Theory*, vol. 69, no. 3, pp. 1883–1898, 2023.

[16] D. Guo, Y. Wu, S. Shamai, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2371–2385, 2011.

[17] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*. Springer Science & Business Media, 2002.