# Estimation Error: Distribution and Pointwise Limits

Luca Barletta*, Alex Dytso†, and Shlomo Shamai (Shitz)**
* Politecnico di Milano, Milano, 20133, Italy. Email: luca.barletta@polimi.it
† Qualcomm Flarion Technologies, Bridgewater, NJ 08807, USA. Email: odytso2@gmail.com
** Technion – Israel Institute of Technology, Haifa 32000, Israel. E-mail: sshlomo@ee.technion.ac.il

***Abstract*—In this paper, we examine the distribution and convergence properties of the estimation error $W = X - \hat{X}(Y)$, where $\hat{X}(Y)$ is the Bayesian estimator of a random variable $X$ from a noisy observation $Y = X + \sigma Z$ where $\sigma$ is the parameter indicating the strength of noise $Z$. Using the conditional expectation framework (that is, $\hat{X}(Y)$ is the conditional mean), we define the normalized error $\mathcal{E}_\sigma = \frac{W}{\sigma}$ and explore its properties.**

**Specifically, in the first part of the paper, we characterize the probability density function of $W$ and $\mathcal{E}_\sigma$. Along the way, we also find conditions for the existence of the inverse functions for the conditional expectations. In the second part, we study pointwise (i.e., almost sure) convergence of $\mathcal{E}_\sigma$ as $\sigma \to 0$ under various assumptions about the noise and the underlying distributions. Our results extend some of the previous limits of $\mathcal{E}_\sigma$ as $\sigma \to 0$ studied under the $L^2$ convergence, known as the *MMSE dimension*, to the pointwise case.**

## I. Introduction

Consider a setting where we seek to estimate a random variable $X$ from a noisy observation $Y$. One accepted way to compute the error is to assess the difference[1] between $X$ and the estimate $\hat{X}(Y)$ (e.g., the Bayesian estimator):

$$W = X - \hat{X}(Y). \tag{1}$$

The 'quality' of the estimate is then typically evaluated by computing some absolute $p$-th moment of $W$, of which the second moment, known as the mean squared error, is by far the most widely used. While the moments of $W$ have received attention in the literature, the distribution of $W$ itself has not been thoroughly investigated. The goal of this work is to start making progress in understanding other properties of $W$, such as its distribution and the rate of convergence of $W$ as some noise parameter (e.g., noise variance) goes to zero, or when some signal-to-noise ratio parameter goes to infinity. Practically, understanding the full distribution of the estimation error $W$ provides deeper insight into the reliability and risk of extreme errors in practical systems beyond what average measures like mean squared error can reveal.

Specifically, in this paper, we focus on scenarios where the noise is modeled as additive, that is:

$$Y = X + \sigma Z \tag{2}$$

where $X$ and $Z$ are independent random variables and $\sigma > 0$ is a noise parameter. Moreover, as the estimator, we take the conditional expectation:

$$\hat{X}(Y) = \mathbb{E}[X|Y] \tag{3}$$

which is an optimal Bayesian estimator under the second moment criterion. We also define the normalized error as:

$$\mathcal{E}_\sigma = \frac{X - \mathbb{E}[X|Y]}{\sigma} = \frac{W}{\sigma} \tag{4}$$

and seek to study the almost sure (a.s.) convergence of $\mathcal{E}_\sigma$ as $\sigma \to 0$.

### A. Definitions and Notation

Throughout the paper, the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is fixed. The density of a random variable $X$ whose distribution is absolutely continuous with respect to Lebesgue measure is denoted by $f_X$. $\phi_\sigma$ denotes the density of Gaussian with zero mean and variance $\sigma^2$. If measure $\mu$ is absolutely continuous with respect to $\lambda$, then it is denoted as $\mu \ll \lambda$. Similarly, $\mu \perp \nu$ denotes that $\mu$ and $\nu$ are mutually singular. $L^1$ denotes the collection of all random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with finite first absolute moment. The conditional expectation of $X$ given $Y$ is defined as

$$\mathbb{E}[X|Y = y] = \int x \mathrm{d}P_{X|Y=y}(x) \tag{5}$$

where $P_{X|Y=y}(x)$ is the conditional distribution. All logarithms are base e.

### B. Paper Outline and Contributions

1) Section II studies the distribution of $W$ and $\mathcal{E}_\sigma$ and shows:
   - Section II-A, Proposition 1, derives sufficient conditions for $y \mapsto \mathbb{E}[X|Y = y]$ to have a functional inverse, which is needed to study the distribution of $W$. In particular, it is shown that if certain conditions hold, most notably $Z$ has a log-concave distribution, then the inverse exists.
   - Section II-B, Theorem 1, characterizes the density function of $W$ and evaluates it for some examples.
2) Section III studies the a.s. convergence of $\mathcal{E}_\sigma$ under various assumptions on the distribution of $X$ and distribution of $Z$.

The rest of this section is dedicated to literature review. Finally, some of the proofs are omitted due to length constraints and can be found in [**?**].

The work of Shamai was supported by the German Research Foundation (DFG) via the German-Israeli Project Cooperation (DIP), under Project SH 1937/1-1.

[1] Other errors are also common; see for example [**?**].

### C. Related Literature

Conditional expectation plays an important role in our discussion. Monotonicity of conditional expectation under rather general settings has been shown in [?], [?]. The derivatives of the conditional expectation under additive and exponential family models have been considered in [?], [?], [?] where the derivative were shown to be proportional to conditional cumulants. The probability distribution for the Gaussian noise has been previously found in [?]. In [?], in the context of Gaussian noise, $W^2$ has been related to the information density via the so-called pointwise I-MMSE relationship. Other related properties of $W$ and the conditional probability $P_{X|Y}$ have been studied in [?], [?], [?], [?], [?]; see [?] for a comprehensive survey of results and applications.

In [?], authors considered convergence of the second moment of $\mathcal{E}_\sigma$ (i.e., $\mathbb{E}[(\mathcal{E}_\sigma)^2]$), which was termed *MMSE dimension*, and provided a fairly complete characterization of the limit. For example, under suitable regularity conditions of the noise distribution $Z$, if the distribution $P_X = \alpha P_{X_c} + (1-\alpha)P_{X_D}$, $\alpha \in [0,1]$ where $P_{X_c}$ is absolutely continuous distribution with respect to Lebesgue measure and $P_{X_D}$ is a discrete distribution, then $\lim_{\sigma\to 0}\mathbb{E}[(\mathcal{E}_\sigma)^2] = \alpha \mathrm{Var}(Z)$. We will provide the pointwise version of this result. Some of the limit theorems shown in [?] would be useful in our analysis too. To find limits, we also borrow techniques from [?].

## II. On the Distribution of the Estimation Error

In this section, we derive the distribution of $X - \mathbb{E}[X|Y]$. To fully derive the result, we need to characterize conditions under which the functional inverse $y \mapsto \mathbb{E}[X|Y=y]$ exists. In addition, we need conditions for this inverse to be differentiable.

### A. On the Inverse Function of the Conditional Expectation

In this section, we study the inverse function of the conditional expectation. For ease of notation, we let $\hat{X}(y) = \mathbb{E}[X|Y=y]$, and let $\hat{X}^{-1}$ denote the functional inverse provided that it exists. The next proposition gives a sufficient condition for the existence of the inverse.

**Proposition 1.** *Suppose that*

- *$X$ is non-degenerate and $X \in L^1$; and*
- *the noise density function can be written as*

$$f_Z(z) = \begin{cases} e^{\psi(z)} & z \in I \\ 0 & o.w. \end{cases} \quad (6)$$

  *where $I$ is an open interval (possibly infinite) and $z \mapsto \psi(z)$ is strictly concave on $I$[2]; and*
- *there exists a function $\theta(X) \in L^1$ such that for all $y \in \mathbb{R}$*

$$|X\psi'(y-X)\exp(\psi(y-X))| \le \theta(X) \text{ a.s.} \quad (7)$$

*Then, $\hat{X}^{-1}$ exists and is differentiable.*

*Proof:* Under the first two assumptions, by slightly modifying the argument of [?, Prop .1] – the modification involves assuming that $\psi$ is strictly concave instead of concave – the conditional expectation $y \mapsto \hat{X}(y)$ is an increasing function instead of just non-decreasing. Since increasing functions have proper functional inverses, the first part of the conclusion follows.

Next, note that

$$\hat{X}(y) = \frac{\mathbb{E}\left[X\exp(\psi(y-X))\right]}{\mathbb{E}\left[\exp(\psi(y-X))\right]} \quad (8)$$

which, by employing Leibniz integral rule to numerator and denominator, is differentiable provided that

$$|X\psi'(y-X)\exp(\psi(y-X))| \le \theta_1(X) \text{ a.s.} \quad (9)$$

$$|\psi'(y-X)\exp(\psi(y-X))| \le \theta_2(X) \text{ a.s.}\ , \quad (10)$$

for some $L^1$ functions $\theta_1$ and $\theta_2$. The proof is concluded by noting that (9) subsumes (10) resulting in condition (7). ∎

The condition of Proposition 1 is only sufficient. In particular, in the case where $Z$ is Gaussian, the existence of the inverse follows from the fact that the derivative of $\hat{X}(y)$ for non-degenerate $X$ is strictly positive since [?], [?]

$$\sigma^2 \frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}[X|Y=y] = \mathsf{Var}(X|Y=y), \quad y \in \mathbb{R}, \quad (11)$$

which holds for all distributions on $X$.

**Example.** Suppose that $X$ and $Z$ are two independent standard Gaussian random variables. Then, the conditional expectation is given by

$$\hat{X}(y) = \frac{1}{1+\sigma^2}y, \quad y \in \mathbb{R}, \quad (12)$$

and the inverse is given by

$$\hat{X}^{-1}(x) = (1+\sigma^2)x, \quad x \in \mathbb{R}. \quad (13)$$

**Example.** Suppose that $X$ is distributed according to $P_X(1) = p = 1 - P_X(-1)$. Then, the conditional expectation is given by

$$\mathbb{E}[X|Y=y] = \tanh\left(\frac{y}{\sigma^2} + \frac{1}{2}\log\left(\frac{p}{1-p}\right)\right),\ y \in \mathbb{R}, \quad (14)$$

and the inverse is given by

$$\hat{X}^{-1}(x) = \frac{\sigma^2}{2}\log\left(\frac{1+x}{1-x}\frac{1-p}{p}\right), \quad x \in (-1,1). \quad (15)$$

### B. Characterization of the Estimation Error Distribution

The next theorem provides an expression for the probability density function (pdf) of the estimation error $X - g(Y)$ where $g$ is some estimator of $X$.

**Theorem 1.** *Let $W = X - g(Y)$ where $g$ has a well-defined functional inverse and where $\mathcal{R}_g$ denotes the range of the function $g$. Then, for $w \in \mathbb{R}$*

$$f_W(w) = \frac{1}{\sigma}\mathbb{E}\Bigg[f_Z\left(\frac{g^{-1}(X-w)-X}{\sigma}\right)\left|\frac{\mathrm{d}g^{-1}(X-w)}{\mathrm{d}w}\right| \cdot \mathbb{1}_{\mathcal{R}_g}(w-X)\Bigg]. \quad (16)$$

[2]This property is known as log-concavity if $I = \mathbb{R}$.

*Consequently, if $g(Y) = \hat{X}(Y)$, for $w \in \mathbb{R}$*

$$f_W(w) = \frac{1}{\sigma}\mathbb{E}\Bigg[f_Z\left(\frac{\hat{X}^{-1}\left(X-w\right)-X}{\sigma}\right)\left|\frac{\mathrm{d}\hat{X}^{-1}(X-w)}{\mathrm{d}w}\right| \cdot \mathbb{1}_{\mathcal{R}_{\hat{X}}}\left(w-X\right)\Bigg]. \tag{17}$$

*Proof:* Let $W = X - g(X+N) = X+U$, where $N = \sigma Z$; then, by using the formula for the sum of correlated variables, we have that

$$f_W(w) = \mathbb{E}\left[f_{U|X}(w-X|X)\right]. \tag{18}$$

To characterize $f_{U|X}(t|x)$, note that given $X = x$

$$U = -g(x+N) = -g(N_x) \tag{19}$$

where $N_x = N + x$. Therefore,

$$\begin{aligned} f_{U|X}(t|x) &= f_{-g(N_x)}(t) & (20)\\ &= f_{N_x}(g^{-1}(-t))\left|\frac{\mathrm{d}}{\mathrm{d}t}g^{-1}(-t)\right|\mathbb{1}_{\mathcal{R}_g}(t) & (21)\\ &= f_{\sigma Z}\left(g^{-1}(-t)-x\right)\left|\frac{\mathrm{d}}{\mathrm{d}t}g^{-1}(-t)\right|\mathbb{1}_{\mathcal{R}_g}(t) & (22)\\ &= \frac{1}{\sigma}f_Z\left(\frac{g^{-1}(-t)-x}{\sigma}\right)\left|\frac{\mathrm{d}}{\mathrm{d}t}g^{-1}(-t)\right|\mathbb{1}_{\mathcal{R}_g}(t), & (23)\end{aligned}$$

where in (21) we have used a change of variable formula. Inserting (23) into (18) concludes the proof of (16). ■

**Remark 1.** Both of the expressions in Theorem 1 can be further simplified or rewritten. In particular, the expression in (16) can be rewritten by using $\frac{\mathrm{d}}{\mathrm{d}t}g^{-1}(t) = \frac{1}{g'(g^{-1}(t))}$. For example, when $Z$ is standard Gaussian, by using (11) we have that

$$\left|\frac{\mathrm{d}\hat{X}^{-1}(w)}{\mathrm{d}w}\right| = \frac{\sigma^2}{\mathsf{Var}\left(X|Y=\hat{X}^{-1}(w)\right)}, \tag{24}$$

which leads to: for $w \in \mathbb{R}$

$$f_W(w) = \sigma^2\mathbb{E}\left[\frac{\phi_\sigma\left(\hat{X}^{-1}(X-w)-X\right)}{\mathsf{Var}\left(X|Y=\hat{X}^{-1}(X-w)\right)}\mathbb{1}_{\mathcal{R}_{\hat{X}}}\left(w-X\right)\right]. \tag{25}$$

We note that if one seeks to find the distribution of $\mathcal{E}_\sigma = \frac{W}{\sigma}$ instead of $W$, then the transformation $f_{\mathcal{E}_\sigma}(w) = \sigma f_W(\sigma w)$, $w \in \mathbb{R}$ can be used.

**Example.** Suppose that $X$ is distributed according to $P_X(1) = 1 - P_X(-1) = p \in (0,1)$. The conditional expectation of the random variable is given in (14), and the inverse is given in (15). The derivative of the inverse is given by $\frac{\mathrm{d}}{\mathrm{d}x}\hat{X}^{-1}(x) = \frac{\sigma^2}{1-x^2}, x \in \mathcal{D}_{\hat{X}}$ where $\mathcal{D}_{\hat{X}} = (-1,1)$. Therefore, by using Theorem 1, the distribution of the error is given by

$$\begin{aligned} f_{\mathcal{E}_\sigma}(w) = {} & \phi_\sigma\left(\frac{\sigma^2}{2}\log\left(\frac{2-\sigma w}{\sigma w}\frac{1-p}{p}\right)-1\right)\frac{\sigma^3 p}{1-(1-\sigma w)^2}\\ & \cdot \mathbb{1}_{(0,2)}(\sigma w)\\ & + \phi_\sigma\left(\frac{\sigma^2}{2}\log\left(\frac{-\sigma w}{2+\sigma w}\frac{1-p}{p}\right)+1\right)\frac{\sigma^3(1-p)}{1-(1+\sigma w)^2}\\ & \cdot \mathbb{1}_{(-2,0)}(\sigma w), \quad w \in \mathbb{R}, \end{aligned} \tag{26}$$

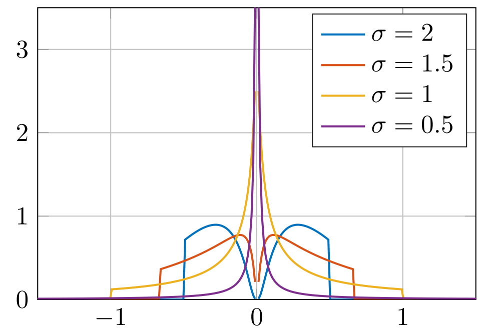


Fig. 1: Density of $\mathcal{E}_\sigma$ in (27).

which in the case of $p = 1/2$ reduces to

$$f_{\mathcal{E}_\sigma}(w) = \frac{1}{2}\phi_\sigma\left(\frac{\sigma^2}{2}\log\left(\frac{2-|\sigma w|}{|\sigma w|}\right)-1\right)\frac{\sigma^3}{1-(1-|\sigma w|)^2} \cdot \mathbb{1}_{(-2,2)}(\sigma w), \quad w \in \mathbb{R}. \tag{27}$$

Figure 1 displays the density in (27).

## III. On $\lim_{\sigma\to 0}\mathcal{E}_\sigma = \mathcal{E}_0$

In this section, we find the distribution of $\mathcal{E}_0$ under various assumption on the distributions of $X$ and $Z$. Table I summarizes the results.

### A. Absolutely Continuous Distributions

In this section, we assume that the distribution of $X$ has a density with respect to Lebesgue measure.

**Theorem 2.** *Suppose that*

- *the density $f_X$ is continuous and bounded; and*
- *$Z \in L^1$.*

*Then,*

$$\lim_{\sigma\to 0}\mathcal{E}_\sigma = \mathbb{E}[Z] - Z \qquad a.s. \tag{28}$$

*Proof:* First observe that

$$\begin{aligned}\mathcal{E}_\sigma &= \frac{X-\mathbb{E}[X|Y]}{\sigma} = \frac{X-Y+Y-\mathbb{E}[X|Y]}{\sigma} & (29)\\ &= -Z + \mathbb{E}[Z|Y]. & (30)\end{aligned}$$

Next, consider for any $x, z \in \mathbb{R}$:

$$\begin{aligned}\lim_{\sigma\to 0}\mathbb{E}[Z|Y=x+\sigma z] &= \lim_{\sigma\to 0}\frac{\mathbb{E}[Zf_X(x+\sigma z-\sigma Z)]}{\mathbb{E}[f_X(x+\sigma z-\sigma Z)]} & (31)\\ &= \frac{\mathbb{E}[Z\lim_{\sigma\to 0}f_X(x+\sigma z-\sigma Z)]}{\mathbb{E}[\lim_{\sigma\to 0}f_X(x+\sigma z-\sigma Z)]} & (32)\\ &= \frac{\mathbb{E}[Z]f_X(x)}{f_X(x)}, & (33)\end{aligned}$$

TABLE I: The expression for $\lim_{\sigma\to 0}\mathcal{E}_\sigma = \mathcal{E}_0$ under various assumptions.

| Assumption on $X$ | Assumption on $Z$ | $\lim_{\sigma\to 0}\mathcal{E}_\sigma = \mathcal{E}_0$ |
|---|---|---|
| discrete distribution | $f_Z$ bounded and $o(\lvert z\rvert^{-1})$ | 0 |
| bounded and continuous density | $Z \in L^1$ | $\mathbb{E}[Z] - Z$ |
| absolutely cont. distribution | $f_Z$ bounded and $O(\lvert z\rvert^{\alpha})$ for some $\alpha > 2$ | |
| $X = \mathbb{1}_{\{U=1\}}X_D + \mathbb{1}_{\{U=2\}}X_c$<br>$U \perp X_D \perp X_c$<br>$P_{X_D}$ − discrete, $P_{X_c}$ − abs. continuous<br>$U \in \{1,2\}$ | Doob's random variable (see Definition 1) | $\mathbb{1}_{\{U=2\}}(\mathbb{E}[Z] - Z)$ |

where (32) follows from boundedness of $f_X$, from $Z \in L^1(\Omega)$, and from dominated convergence theorem; (33) follows from continuity of $f_X$. Since $f_X(X) > 0$ a.s., we have $\lim_{\sigma\to 0}\mathbb{E}[Z|Y] = \mathbb{E}[Z]$ a.s. which, combined with (30), gives the claimed result. ∎

We now put more assumptions on the noise $Z$ while keeping minimal assumption on $X$.

**Theorem 3.** *Suppose that $Z$ has a bounded density and such that for some $\alpha > 2$,*

$$f_Z(z) = O(|z|^{-\alpha}), \qquad |z| \to \infty. \tag{34}$$

*Then, for all $X$ with absolutely continuous distribution*

$$\lim_{\sigma\to 0}\mathcal{E}_\sigma = \mathbb{E}[Z] - Z \qquad a.s. \tag{35}$$

*Proof:* Let us introduce the density of variable $\sigma(Z - z)$:

$$K_\sigma(x) = \frac{1}{\sigma} f_Z\left(\frac{x}{\sigma} + z\right). \tag{36}$$

Notice that

$$f_Y(x + \sigma z) = f_X * K_\sigma(x), \tag{37}$$

where $*$ denotes convolution. Then, by using $f_X \in L^1(\mathbb{R})$, the conditions on $f_Z$, and [**?**, Lemma 8], we can say that

$$\lim_{\sigma\to 0} f_Y(x + \sigma z) = f_X(x) \tag{38}$$

holds for Lebesgue-a.e. $x$. Let us check that the conditions of [**?**, Lemma 8] for the function $K_\sigma$ are satisfied:

- $\int_{\mathbb{R}} K_\sigma(x)\mathrm{d}x = 1$;
- $\sup_{x\in\mathbb{R},\,\sigma>0} \sigma|K_\sigma(x)| = \sup_{x\in\mathbb{R}} |f_Z(x)| < \infty$ because of the boundedness condition on $f_Z$;
- $$\left(\sup_{x\in\mathbb{R},\,\sigma>0} \frac{|x|^{1+\eta}}{\sigma^\eta}|K_\sigma(x)|\right)^{\frac{1}{1+\eta}} \tag{39}$$
$$= \sup_{u\in\mathbb{R}} |u| f_Z(u+z)^{\frac{1}{1+\eta}} \tag{40}$$
$$\leq \sup_{u\in\mathbb{R}} |u| f_Z(u)^{\frac{1}{1+\eta}} + |z| \sup_{u\in\mathbb{R}} f_Z(u)^{\frac{1}{1+\eta}} \tag{41}$$
$$< \infty \tag{42}$$
because of the boundedness condition on $f_Z$ and because $f_Z(u) = O(|u|^{-2-\eta})$ for some $\eta > 0$.

Next, introduce the function

$$G_\sigma(x) = \left(\frac{x}{\sigma} + z\right)\frac{1}{\sigma} f_Z\left(\frac{x}{\sigma} + z\right), \tag{43}$$

and notice that

$$\mathbb{E}[Z f_X(x + \sigma z - \sigma Z)] = f_X * G_\sigma(x). \tag{44}$$

Then, by using $f_X \in L^1(\mathbb{R})$, the conditions on $f_Z$, and [**?**, Lemma 8], we can say that

$$\lim_{\sigma\to 0}\mathbb{E}[Z f_X(x + \sigma z - \sigma Z)] = \mathbb{E}[Z] f_X(x) \tag{45}$$

holds for Lebesgue-a.e. $x$. Let us check that the conditions of [**?**, Lemma 8] for the function $G_\sigma$ are satisfied:

- $\int_{\mathbb{R}} G_\sigma(x)\mathrm{d}x = \mathbb{E}[Z]$;
- $\sup_{x\in\mathbb{R},\,\sigma>0} \sigma|G_\sigma(x)| = \sup_{u\in\mathbb{R}} |u| f_Z(u) < \infty$ because of the boundedness condition on $f_Z$ and because $f_Z(u) = O(|u|^{-2-\eta})$ for some $\eta > 0$;
- $$\left(\sup_{x\in\mathbb{R},\,\sigma>0} \frac{|x|^{1+\eta}}{\sigma^\eta}|G_\sigma(x)|\right)^{\frac{1}{1+\eta}} \tag{46}$$
$$= \sup_{u\in\mathbb{R}} |u|\, |u+z|^{\frac{1}{1+\eta}} f_Z(u+z)^{\frac{1}{1+\eta}} \tag{47}$$
$$\leq \sup_{u\in\mathbb{R}} |u|^{1+\frac{1}{1+\eta}} f_Z(u)^{\frac{1}{1+\eta}} + |z| \sup_{u\in\mathbb{R}} |u|^{\frac{1}{1+\eta}} f_Z(u)^{\frac{1}{1+\eta}}$$
$$< \infty \tag{48}$$
because of the boundedness condition on $f_Z$ and because $f_Z(u) = O(|u|^{-2-\eta})$ for some $\eta > 0$.

Using the results (38) and (45), we have

$$\lim_{\sigma\to 0}\mathbb{E}[Z|Y = x + \sigma z] = \lim_{\sigma\to 0}\frac{\mathbb{E}[Z f_X(x + \sigma z - \sigma Z)]}{\mathbb{E}[f_X(x + \sigma z - \sigma Z)]} \tag{49}$$
$$= \mathbb{E}[Z] \quad \text{a.s.} \tag{50}$$

∎

### *B. Discrete Distributions*

We now consider discrete distributions. The next theorem makes the statement that *discrete random variables are 'easy' to estimate* more concrete.

**Theorem 4.** *If $X$ is discrete (finitely or countable infinitely valued), then*

$$\lim_{\sigma\to 0}\mathcal{E}_\sigma = 0 \quad a.s. \tag{51}$$

*holds for all $Z$ with bounded density and such that*

$$f_Z(z) = o(|z|^{-1}), \qquad |z| \to \infty. \tag{52}$$

*Proof:* Start from

$$\mathcal{E}_\sigma = \mathbb{E}[Z|Y] - Z. \tag{53}$$

Let $p_i := \Pr(X = x_i)$. For all $z$ and $x_i$, define the function

$$g_{z,i}(\sigma) := |z - \mathbb{E}[Z|Y = x_i + \sigma z]|. \tag{54}$$

Then, for all $z \in \mathbb{R}$ and $x_i \in \mathsf{supp}(P_X)$ we have

$$\begin{aligned} g_{z,i}(\sigma) &= \frac{\left|\mathbb{E}\left[\frac{x_i - X}{\sigma} f_Z\left(\frac{x_i - X}{\sigma} + z\right)\right]\right|}{\sum_j p_j f_Z\left(\frac{x_i - x_j}{\sigma} + z\right)} & (55)\\ &\le \frac{\mathbb{E}\left[\frac{|x_i - X|}{\sigma} f_Z\left(\frac{x_i - X}{\sigma} + z\right)\right]}{p_i f_Z(z)}. & (56)\end{aligned}$$

Next, fix some $d > 0$ and note that

$$\begin{aligned} &\frac{|x_i - X|}{\sigma} f_Z\left(\frac{x_i - X}{\sigma} + z\right)\\ &= \frac{|x_i - X|}{\sigma} f_Z\left(\frac{x_i - X}{\sigma} + z\right)\left(\mathbb{1}_{\left\{\frac{|x_i - X|}{\sigma} \le d\right\}} + \mathbb{1}_{\left\{\frac{|x_i - X|}{\sigma} > d\right\}}\right) & (57)\\ &\le dB + \frac{|x_i - X|}{\sigma} f_Z\left(\frac{x_i - X}{\sigma} + z\right)\mathbb{1}_{\left\{\frac{|x_i - X|}{\sigma} > d\right\}} & (58)\\ &\le dB + C, & (59)\end{aligned}$$

where $\sup_z f(z) \le B$ and where $C < \infty$ which exists since $f_Z(u) = o(|u|^{-1})$.

Now combining the bound in (59) with the dominated convergence theorem, we arrive at

$$\lim_{\sigma \to 0} \mathbb{E}\left[\frac{|x_i - X|}{\sigma} f_Z\left(\frac{x_i - X}{\sigma} + z\right)\right] = 0. \tag{60}$$

As a consequence of (56), (60) and $f_Z(Z) > 0$ a.s., we have that $\lim_{\sigma \to 0} |\mathcal{E}_\sigma| = 0$ a.s., and the claim follows. ■

### C. Mixed Random Variables

In the previous two sections, we have considered the case where the input $X$ was discrete or continuous. We now consider the case of mixed random variables and assume that $X$ is composed as follows:

$$X = \mathbb{1}_{\{U=1\}} X_1 + \mathbb{1}_{\{U=2\}} X_2, \tag{61}$$

where $X_i$ is a random variable with distribution $\mu_i$, for $i = 1, 2$, and $U$ is a random variable independent of $X_1, X_2$, taking values on $\{1, 2\}$ with probability $\mathbb{P}[U = 1] = \alpha \in (0, 1)$. Note that the distribution of $X$ is

$$\mu = \alpha \mu_1 + (1 - \alpha)\mu_2. \tag{62}$$

Of particular interest in this section is the case when $\mu_1 \perp \mu_2$, which can include the case when $\mu_1$ is discrete and $\mu_2$ is continuous.

Before proceeding with the general result, we also need to impose regularity conditions on the additive noise. We adopt the same regularity conditions as those used in [**?**] which are fairly general.

**Definition 1.** We refer to a random variable $Z$ as a Doob's random variable if its density satisfies the conditions in [**?**, Thm. 4.1] (see also Lemma 1 in [**?**]).

Loosely speaking, Doob's condition controls the smoothness and the tail behavior of the distribution of $Z$. Many practically relevant random variables, such as the Gaussian, satisfy Doob's conditions.

We begin by proving the following decomposition result.

**Theorem 5.** *Given an input as in* (61) *suppose that*

- *$\mu_1 \perp \mu_2$; and*
- *$Z$ is a Doob's random variable.*

*and for $u \in \{1, 2\}$ define*

$$\begin{aligned} Y_u &= X_u + Z, & (63)\\ \mathcal{E}_{u,\sigma} &:= \frac{X_u - \mathbb{E}[X_u|Y_u]}{\sigma} = -Z + \mathbb{E}[Z|Y_u]. & (64)\end{aligned}$$

*Then, for every $\alpha \in [0, 1]$,*

$$\lim_{\sigma \to 0} |\mathcal{E}_\sigma - \mathbb{1}_{\{U=1\}}\mathcal{E}_{1,\sigma} - \mathbb{1}_{\{U=2\}}\mathcal{E}_{2,\sigma}| = 0 \qquad a.s. \tag{65}$$

*Proof:* See [**?**]. ■

Now, as a corollary of Theorem 5 and previous results, we have the following conclusion.

**Corollary 1.** *Suppose that*

- *$X$ in* (61) *be such that $X_1$ is discrete and $X_2$ has an absolutely continuous distribution; and*
- *$Z$ is a Doob's random variable.*

*Then, for every $\alpha \in [0, 1]$,*

$$\lim_{\sigma \to 0} \mathcal{E}_\sigma = \mathbb{1}_{\{U=2\}}\mathcal{E}_{2,\sigma} = \mathbb{1}_{\{U=2\}}(\mathbb{E}[Z] - Z) \ a.s. \tag{66}$$

We conclude this section by presenting a decomposition theorem for two measures that are absolutely continuous.

**Theorem 6.** *Given an input as in* (61) *suppose that*

- *$\mu_1 \ll \mu_2$; and*
- *$Z$ is a Doob's random variable.*

*Then, for every $\alpha \in [0, 1]$,*

$$\lim_{\sigma \to 0} |\mathcal{E}_\sigma - \mathbb{1}_{\{U=1\}}\mathcal{E}_{1,\sigma} - \mathbb{1}_{\{U=2\}}\mathcal{E}_{2,\sigma}| = 0 \qquad a.s. \tag{67}$$

*Proof:* See [**?**]. ■

As before, Theorem 6 is only a decomposition result and limits of $\mathcal{E}_{2,\sigma}$, if they exists, would have to be found separately.

## IV. Conclusion

This work departed from the standard moment-based treatment of the estimation error and instead focused on its distributional properties and the pointwise convergence. The main focus was on additive noise channels. The paper derived the structure of the probability density function of the estimation error. Additionally, the pointwise convergence of the estimation error in the low-noise regime was characterized for a fairly general setting.

As a future direction, it would be interesting to characterize similar pointwise limits for the information density and to find an equivalent limit for the information dimension [**?**]. It will also be interesting to connect the limits in this work with the pointwise I-MMSE relationship [**?**]. In this work, the conditional mean was used as the estimator, but it would also be interesting to consider other estimators, such as the conditional median [**?**].