On the Distribution of the Conditional Mean Estimator in Gaussian Noise

Alex Dytso*, H. Vincent Poor† and Shlomo Shamai (Shitz)‡

*Department of Electrical and Computer Engineering, New Jersey Institute of Technology

†Department of Electrical Engineering, Princeton University

‡Department of Electrical Engineering, Technion – Israel Institute of Technology

Email: alexdytso@njit.edu*, poor@princeton.edu† and sshlomo@ee.technion.ac.il‡

***Abstract*—Consider the conditional mean estimator of the random variable X from the noisy observation $Y = X + N$ where N is zero mean Gaussian with variance σ^2 (i.e., $\mathbb{E}[X|Y]$). This work characterizes the probability distribution of $\mathbb{E}[X|Y]$. As part of the proof, several new identities and results are shown. For example, it is shown that the k-th derivative of the conditional expectation is proportional to the $(k+1)$-th conditional cumulant. It is also shown that the compositional inverse of the conditional expectation is well-defined and is characterized in terms of a power series.**

***Index Terms*—Conditional mean estimator, Gaussian Noise.**

A full version of this paper is accessible at [1].

I. Introduction

Consider a setting in which a random variable X is observed through an additive Gaussian noise channel

$$Y = X + N, \tag{1}$$

where $N \sim \mathcal{N}(0, \sigma^2)$ and is independent of X. Throughout the paper, σ^2 is assumed to be positive, and we make no assumptions about X other than $\mathbb{E}[X^2] < \infty$.

It is well-known that the optimal minimum mean squared error (MMSE) estimator of X from the observation Y is given by *the conditional expectation*

$$\hat{X}(Y) = \mathbb{E}[X|Y] = \int X \mathrm{d}P_{X|Y}. \tag{2}$$

The quantity $\mathbb{E}[X|Y]$, which is a random variable, has a wide range of applications in probability, statistics and information theory. The goal of this work is to study the distribution of the conditional expectation under the Gaussian noise setting. The distribution of a given estimator typically contains more information than measures like variance and is more useful. For example, such a distribution can be used to obtain confidence intervals. The difficulty of finding the distribution $\hat{X}(Y)$ lies in the fact that the conditional expectation seldom has a closed-form expression.

The distribution of $\hat{X}(Y)$ is closely related to P_X and $P_{X|Y}$. However, while the distributions P_X or $P_{X|Y}$ can be arbitrary (discrete, continuous or singular), the random variable $\hat{X}(Y)$ is always continuous. This follows from the fact that Y is a continuous random variable, and, as will be shown in what follows, the fact that the function $y \mapsto \hat{X}(y)$ is real-analytic.

This work was supported in part by the U. S. National Science Foundation under Grant CCF-1908308 and by the United States-Israel Binational Science Foundation, under Grant BSF-2018710, and in part by the European Union's Horizon 2020 Research And Innovation Programme, grant agreement no. 694630.

Our starting place for finding the distribution of $\hat{X}(Y)$ is the following well-known change of variable formulas: for the random variable V with the cumulative distribution function (cdf) F_V and the probability density function (pdf) f_V, let $W = g(V)$, then

$$F_W(w) = F_V(g^{-1}(w)), \tag{3}$$

$$f_W(w) = f_V(g^{-1}(w)) \left| \frac{\mathrm{d}}{\mathrm{d}w} g^{-1}(w) \right|, \tag{4}$$

where g^{-1} the inverse of g. Assuming that these transformation apply to our setting, we arrive at

$$F_{\hat{X}}(x) = F_Y(\hat{X}^{-1}(x)), \tag{5}$$

$$f_{\hat{X}}(x) = f_Y\left(\hat{X}^{-1}(x)\right) \left| \frac{\mathrm{d}}{\mathrm{d}x} \hat{X}^{-1}(x) \right|, \tag{6}$$

where $\hat{X}^{-1}(x)$ is the inverse of $\hat{X}(y) = \mathbb{E}[X|Y = y]$, and F_Y and f_Y are the cdf and the pdf of Y, respectively. It is important to note that the function $y \mapsto \hat{X}(y)$ (and the inverse $\hat{X}^{-1}$) depends on the joint distribution P_{XY}.

The program for the rest of the paper is the following. First, we need to demonstrate that $\hat{X}^{-1}(x)$ exists and is differentiable. Second, we need to provide a non-trivial expression for the inverse $\hat{X}^{-1}(x)$. This will enable the application of formulas in (5) and (6).

The question of finding the distribution of $\mathbb{E}[X|Y]$ is akin to *the information spectrum method* [2] where the objective is to find the distribution of the *information density* $\iota_{P_{XY}}(x; y) = \log \frac{\mathrm{d}P_{XY}}{\mathrm{d}(P_X \otimes P_Y)}(x, y)$. Indeed, by using the identity

$$\mathbb{E}[X|Y] = y + \sigma^2 \frac{\mathrm{d}}{\mathrm{d}y} \log f_Y(y), \tag{7}$$

the derivative of the information density can be expressed as

$$\frac{\mathrm{d}}{\mathrm{d}y} \iota_{P_{XY}}(x; y) = \frac{x - \mathbb{E}[X|Y = y]}{\sigma^2}. \tag{8}$$

In this paper, however, we do not pursue this connection.

The identity in (7) will be used several times throughout the paper and was first derived by Robbins in [3] where he credits Maurice Tweedie for the derivation. The vector version of the identity in (7) was derived by Esposito in [4]. Therefore, throughout this paper, we refer to the identity in (7) as the Tweedie-Robbins-Esposito identity or TRE for short.

Contribution: The contribution and the outline of the paper are as follows:

1) Section II provides two examples for which the inverse and the distribution of the conditional expectation can be found in closed-form; and

2) Section III discusses connections between the conditional expectation and the conditional cumulants. In particular, it is shown that the k-th derivative of the conditional expectation is proportional to the $(k+1)$-th conditional cumulant;
3) Section IV finds a power series expansion for the conditional expectation. Moreover, the inverse of the conditional expectation is shown to be well-defined and is characterized in terms of a power series;
4) Section V combines the results of Section III and Section IV and provides the characterization of the distribution of $\hat{X}(Y)$. The distribution of $\hat{X}(Y)$ is shown to depend on the joint distribution P_{XY} only through the marginal f_Y. Finally, several numerical examples are shown; and
5) Section VI concludes the paper.

Notation: Deterministic quantities are denoted by lowercase letters and random variables are denoted by uppercase letters. The (n,k)-th *partial Bell polynomial* is denoted by $\mathsf{B}_{n,k}(x_1,\dots,x_{n-k+1})$. The pdf and cdf of standard Gaussian distribution is denoted by $\phi(x)$ and $\Phi(x)$, respectively.

II. Examples with Closed-Form Expressions for the Distribution

In this section, we consider two example in which the distribution of $\hat{X}(y)$ can be found in closed-form.

Example. Suppose that X is a standard Gaussian random variable. Then, the conditional expectation is given by

$$\hat{X}(y) = \frac{1}{1+\sigma^2} y,\ y \in \mathbb{R}, \tag{9}$$

and the inverse is given by

$$\hat{X}^{-1}(x) = (1+\sigma^2)x,\ x \in \mathbb{R}. \tag{10}$$

Then, using (6), for $x \in \mathbb{R}$

$$f_{\hat{X}}(x) = f_Y\left((1+\sigma^2)x\right)\left(1+\sigma^2\right) = \frac{1}{\sqrt{2\pi\frac{1}{1+\sigma^2}}} \mathrm{e}^{-\frac{x^2}{2\frac{1}{1+\sigma^2}}}. \tag{11}$$

In other words, $\hat{X}$ is Gaussian with variance $\frac{1}{1+\sigma^2}$.

Example. Suppose that X uniformly distributed on $\{-1,1\}$. Then,

$$\hat{X}(y) = \tanh\left(\frac{y}{\sigma^2}\right),\ y \in \mathbb{R} \tag{12}$$

and the inverse is given by

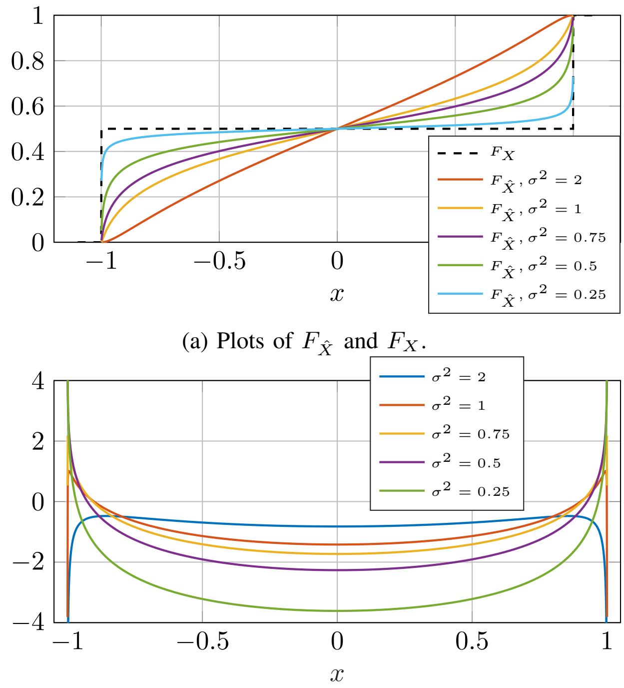
$$\hat{X}^{-1}(x) = \frac{\sigma^2}{2}\log\left(\frac{1+x}{1-x}\right), x \in (-1,1). \tag{13}$$

Then, from $\frac{\mathrm{d}}{\mathrm{d}x}\hat{X}^{-1}(x) = \frac{\sigma^2}{1-x^2}$, for $x \in (-1,1)$ we have that

$$F_{\hat{X}}(x) = f_Y(\hat{X}^{-1}(x)), \tag{14}$$

$$f_{\hat{X}}(x) = f_Y\left(\hat{X}^{-1}(x)\right)\frac{\sigma^2}{1-x^2}, \tag{15}$$

where $F_Y(y) = \frac{1}{2}\Phi\left(\frac{y-1}{\sigma}\right) + \frac{1}{2}\Phi\left(\frac{y+1}{\sigma}\right)$ and $f_Y(y) = \frac{1}{2\sigma}\phi(\frac{y-1}{\sigma}) + \frac{1}{2\sigma}\phi(\frac{y+1}{\sigma})$. The plot of the cdf of $\hat{X}$ is given in Fig. 1 and is compared to the cdf of X. Because as $\sigma \to 0$ the pdf of $\hat{X}$ starts to concentrate on 1 and -1, it is more convenience to plot $\log f_{\hat{X}}$ instead of $f_{\hat{X}}$. The plots of the log of the pdf of $\hat{X}$ is given in Fig. 1b.



(a) Plots of $F_{\hat{X}}$ and F_X.

(b) Plots of $\log f_{\hat{X}}$.

Fig. 1: The cdf and the pdf of $\hat{X}$.

In the above examples, we had closed-form expressions for the conditional expectation. Most likely, there other examples for which closed-form expressions can be found. However, in the general, we do not have such a luxury. In the next two sections, we develop auxiliary results needed to find the inverse and the distribution of the conditional expectation.

III. Connections between the Conditional Cumulants and the Conditional Expectation

In this section, we establish an identity that connects the conditional expectation and the conditional cumulants. To that end, recall that the *conditional cumulant generating function* is given by

$$K_X(t|Y=y) = \log\left(\mathbb{E}[\mathrm{e}^{tX}|Y=y]\right),\ y \in \mathbb{R}, t \in \mathbb{R}, \tag{16}$$

and the k-th *conditional cumulant* is given by

$$\kappa_{X|Y=y}(k) = \frac{\mathrm{d}^k}{\mathrm{d}t^k}K_X(t|Y=y)\Big|_{t=0}, k \in \mathbb{N}. \tag{17}$$

Our starting place is the following general derivative identity shown in [5, Thm. 1].

Lemma 1. *Let $U \to X \to Y$ form a Markov chain. Then,*

$$\sigma^2 \frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}[U|Y=y] = \mathrm{Cov}(X,U|Y=y),\ y \in \mathbb{R}, \tag{18}$$

where the conditional covariance is given by $\mathrm{Cov}(X,U|Y=y) = \mathbb{E}[XU|Y=y] - \mathbb{E}[X|Y=y]\mathbb{E}[U|Y=y]$.

Using the identity in (18) we arrive at the following relationship between the conditional cumulant generating function and the conditional expectation.

Theorem 1. *For $t, y \in \mathbb{R}$ and $k \in \mathbb{N} \cup \{0\}$*

$$\frac{\mathrm{d}^{k+1}}{\mathrm{d}t^{k+1}} K_X(t|Y=y) = \sigma^{2(k+1)} \frac{\mathrm{d}^{k+1}}{\mathrm{d}y^{k+1}} K_X(t|Y=y) + \sigma^{2k} \frac{\mathrm{d}^k}{\mathrm{d}y^k} \mathbb{E}[X|Y=y]. \quad (19)$$

Proof: First, consider the case of $k=0$. By setting $U = \mathrm{e}^{tX}$, $t \in \mathbb{R}$ in (18) we arrive at

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}y} K_X(t|Y=y) &= \frac{\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}[\mathrm{e}^{tX}|Y=y]}{\mathbb{E}[\mathrm{e}^{tX}|Y=y]} && (20)\\ &= \frac{1}{\sigma^2} \frac{\mathbb{E}[X\mathrm{e}^{tX}|Y=y] - \mathbb{E}[\mathrm{e}^{tX}|Y=y]\mathbb{E}[X|Y=y]}{\mathbb{E}[\mathrm{e}^{tX}|Y=y]} && (21)\\ &= \frac{1}{\sigma^2} \frac{\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\mathrm{e}^{tX}|Y=y] - \mathbb{E}[\mathrm{e}^{tX}|Y=y]\mathbb{E}[X|Y=y]}{\mathbb{E}[\mathrm{e}^{tX}|Y=y]} && (22)\\ &= \frac{1}{\sigma^2}\frac{\mathrm{d}}{\mathrm{d}t} \log(\mathbb{E}[\mathrm{e}^{tX}|Y=y]) - \mathbb{E}[X|Y=y] && (23)\\ &= \frac{1}{\sigma^2}\left(\frac{\mathrm{d}}{\mathrm{d}t} K_X(t|Y=y) - \mathbb{E}[X|Y=y]\right). && (24) \end{aligned}$$

The rest of the proof follows by using (24) together with a simple induction argument. ∎

As a corollary of the above result, we arrive at the following relationship between the conditional cumulants and the conditional expectation.

Corollary 1. *For $y \in \mathbb{R}$ and $k \in \mathbb{N} \cup \{0\}$*

$$\kappa_{X|Y=y}(k+1) = \sigma^{2k} \frac{\mathrm{d}^k}{\mathrm{d}y^k} \mathbb{E}[X|Y=y]. \quad (25)$$

Proof: The proof follows from an observation that $\frac{\mathrm{d}^{k+1}}{\mathrm{d}y^{k+1}} K_X(t|Y=y)|_{t=0} = 0$. ∎

Remark 1. It is well-known that the cumulants and the moments of a random variable U have a one-to-one correspondence with the inverse relationship given by

$$\kappa_U(k) = \sum_{m=1}^{k} c_m \mathsf{B}_{k,m}(\mu_1, \ldots, \mu_{k-m+1}), \quad (26)$$

where $\mu_m = \mathbb{E}[U^m]$ [6, Example 11.4]. As shown in an extended version of this work [1], combining (26) together with (18) leads to an alternative proof of (25).

Remark 2. From Corollary 1 we make the following two observations:

- For $k \in \mathbb{N} \cup \{0\}$
$$\sigma^2 \frac{\mathrm{d}}{\mathrm{d}y} \kappa_{X|Y=y}(k+1) = \kappa_{X|Y=y}(k+2); \text{ and} \quad (27)$$
- Using the TRE identity in (7), we arrive at the representation of cumulants in terms of f_Y only
$$\kappa_{X|Y=y}(1) = y + \sigma^2 \frac{\mathrm{d}}{\mathrm{d}y} \log f_Y(y), \quad (28a)$$
$$\kappa_{X|Y=y}(2) = \sigma^2 + \sigma^4 \frac{\mathrm{d}^2}{\mathrm{d}y^2} \log f_Y(y), \quad (28b)$$
$$\kappa_{X|Y=y}(k) = \sigma^{2k} \frac{\mathrm{d}^k}{\mathrm{d}y^k} \log f_Y(y), k \geq 3. \quad (28c)$$

In other words, this shows that the conditional cumulants depend on P_{XY} only through the marginal f_Y.

Example. In the case when X is standard Gaussian the conditional expectation $\mathbb{E}[X|Y=y]$ is a linear function of y. Therefore, by using (25), we have that

$$\kappa_{X|Y=y}(1) = \frac{1}{1+\sigma^2} y, \ \kappa_{X|Y=y}(2) = \frac{\sigma^2}{1+\sigma^2}, \quad (29)$$
$$\kappa_{X|Y=y}(k) = 0, \ k \geq 3. \quad (30)$$

Note that this is as expected since $P_{X|Y}$ is Gaussian, and for the Gaussian distribution only the first and the second cumulants are non-zero.

Example. Consider X uniformly distributed on $\{-3, 0, 3\}$. Fig. 2 shows plots of $\kappa_{X|Y=y}$ vs. y for several values of k.

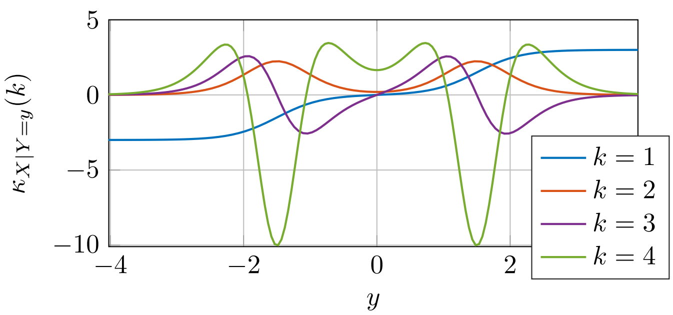


Fig. 2: Plot of $\kappa_{X|Y=y}(k)$ vs. y for $k = 1, 2, 3$ and 4.

We now show that (25) can be used to produce bounds on the rate of growth of the conditional expectation.

Proposition 1. *For $y \in \mathbb{R}$ and $k \in \mathbb{N}$*

$$|\kappa_{X|Y=y}(k)| \leq 2^{k-1} k^k \mathbb{E}[|X|^k|Y=y] \leq a_k |y|^k + b_k, \quad (31)$$

where

$$a_k = k^k 2^{k-1}(2^{\max(\frac{k}{2}-1,1)} + 2), \quad (32)$$
$$b_k = k^k(2^{\max(\frac{k}{2}-1,1)+k} \mathbb{E}^{\frac{k}{2}}[X^2] + \mathbb{E}[|X|^k]). \quad (33)$$

IV. The Power Series and the Inverse of the Conditional Expectation

In this section, we find the power series expansion of the conditional expectation in terms of the conditional cumulants. Furthermore, this power series representation, together with the Lagrange inversion theorem, will lead to a representation of the inverse of the conditional expectation.

A. A Power Series Expansion of $\mathbb{E}[X|Y=y]$

The fact that a power series expansion exists follows from the next result.

Lemma 2. *The functions $y \to \mathbb{E}[X|Y=y]$ is real-analytic.*

Proof: Note that by the TRE identity in (7)

$$\mathbb{E}[X|Y=y] = y + \sigma^2 \frac{\frac{\mathrm{d}}{\mathrm{d}y} f_Y(y)}{f_Y(y)}. \quad (34)$$

Hence, since the ratios and sums of analytic functions are analytic, $\mathbb{E}[X|Y=y]$ is analytic provided that $f_Y(y)$ is analytic. The analyticity of f_Y is a known consequence of convolution with Gaussian measures (see e.g., [7]). ∎

Before studying the Taylor series of the conditional expectation, it is instructive to consider the following example.

Example. For X uniformly distributed on $\{-1, 1\}$

$$\mathbb{E}[X|Y=y] = \tanh\left(\frac{y}{\sigma^2}\right).$$

By using the Taylor series of $\tanh$ around zero, we have that

$$\mathbb{E}[X|Y=y] = \sum_{k=1}^{\infty} \frac{2^{2k}(2^{2k}-1)b_{2k}}{(2k)!}\left(\frac{y}{\sigma^2}\right)^{2k-1}, \ |y| < \frac{\sigma^2\pi}{2},$$

where b_n is the n-the Bernoulli number.

The key observation here is that even in this simple case, the Taylor expansion has a finite radius of convergence. Therefore, in general, we cannot expect to get a power series representation of $\mathbb{E}[X|Y=y]$ that converges for all $\mathbb{R}$ (i.e., the power series with an infinite radius of convergence).

The next result provides a power series representation for the conditional expectation.

Theorem 2. *Fix some $a \in \mathbb{R}$. Then, for every X there exists an $r_{\sigma,a} > 0$ such that*

$$\mathbb{E}[X|Y=y] = \sum_{k=0}^{\infty} \frac{\kappa_{X|Y=a}(k+1)}{k!\sigma^{2k}}(y-a)^k, \ |y-a| \le r_{\sigma,a}. \tag{35}$$

In addition, if $|X| \le A$, then $r_{\sigma,a} \ge \frac{\sigma^2}{2A\mathrm{e}}$.

Remark 3. It is important to note that since the conditional cumulants can be expressed in terms of f_Y and derivatives of f_Y (see (28)), the power series can also be expressed in terms of f_Y only.

B. Inverse of the Conditional Expectation

In this section, we find the inverse of the conditional expectation.

Lemma 3. *Suppose that X is non-constant random variable. Then, $\hat{X}(y)$ has a compositional inverse $\hat{X}^{-1}$. Moreover, the inverse $\hat{X}^{-1}$ is a real-analytic function.*

Proof: By choosing $U = X$ in (18) we arrive at[1]

$$\sigma^2 \frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}[X|Y=y] = \mathrm{Var}(X|Y=y), \ y \in \mathbb{R}. \tag{36}$$

From (36) we have that $\hat{X}(y)$ is a strictly increasing function for non-constant random variables. Therefore, it has a proper inverse. The proof is concluded by using [9, Thm. 1.5.3] which states that the inverse of an analytic function with non-vanishing derivative is also analytic. ■

In order to find the inverse of the conditional expectation we use the power series expansion of the conditional expectation in Theorem 2 and the Lagrange inversion theorem [6].

Theorem 3. *(*Lagrange Inversion Theorem) *The Taylor coefficients of a formal power series $f^{-1}(t) = \sum_{n=1}^{\infty} b_n \frac{t^n}{n!}$, which is the inverse of $f(t) = \sum_{n=1}^{\infty} a_n \frac{t^n}{n!}$, can be expressed as a function of the Taylor coefficients of f in the following manner:*

$$b_1 = \frac{1}{a_1}, \tag{37}$$

$$b_n = b_1^n \sum_{k=1}^{n-1} (-1)^k n^{(k)} \mathsf{B}_{n-1,k}\left(c_1, c_2, \dots, c_{n-k}\right), n \ge 2, \tag{38}$$

where $n^{(k)}$ is the rising factorial and $c_n = \frac{a_{n+1}}{(n+1)a_1}$.

The main result of this section is the following theorem that characterizes the inverse of the conditional expectation.

Theorem 4. *Fix an $a \in \mathbb{R}$. Then, for every non-constant X there exists a $\tau_{\sigma,a} > 0$ such that*

$$\hat{X}^{-1}(x) = a + \sum_{k=1}^{\infty} b_k \frac{\left(x - \hat{X}(a)\right)^k}{k!}, \ |x - \hat{X}(a)| < \tau_{\sigma,a} \tag{39}$$

where

$$b_1 = \frac{\sigma^2}{\kappa_{X|Y=a}(2)}, \tag{40}$$

$$b_n = b_1^n \sum_{k=1}^{n-1} (-1)^k n^{(k)} \mathsf{B}_{n-1,k}\left(c_1, c_2, \dots, c_{n-k}\right), n \ge 2, \tag{41}$$

$$c_k = \frac{\kappa_{X|Y=a}(k+2)}{(k+1)\sigma^{2(k+1)}\kappa_{X|Y=a}(2)}, k \ge 1. \tag{42}$$

Proof: First, since $\hat{X}(y)$ is real-analytic, it has a power-series expansion around $\hat{X}(a)$ with some positive radius of convergence $\tau_{\sigma,a}$. Second, by using (35) we have that

$$f(y) = \hat{X}(y+a) - \hat{X}(a) = \sum_{k=1}^{\infty} \frac{\kappa_{X|Y=a}(k+1)}{k!\sigma^{2k}} y^k = \sum_{k=1}^{\infty} a_k \frac{y^k}{k!}, \tag{43}$$

where $a_k = \frac{\kappa_{X|Y=a}(k+1)}{\sigma^{2k}}$. Therefore, by the Lagrange inversion theorem, we have that

$$f^{-1}(x) = \sum_{k=1}^{\infty} b_k \frac{x^k}{k!}, \tag{44}$$

where the expression for b_k's are given in (38). Next, by noting that $f^{-1}(x) = \hat{X}^{-1}\left(x + \hat{X}(a)\right) - a$, we arrive at

$$\hat{X}^{-1}(x) = a + \sum_{k=1}^{\infty} b_k \frac{\left(x - \hat{X}(a)\right)^k}{k!}. \tag{45}$$

This concludes the proof. ■

Remark 4. We remark that the inverse of the conditional expectation in (39) depends on the joint distribution P_{XY} only through the marginal f_Y. Indeed, by using (28) we have that the coefficients in (42) can be expressed only in terms of f_Y or derivatives of f_Y

$$c_k = \frac{\frac{\mathrm{d}^{k+2}}{\mathrm{d}y^{k+2}} \log f_Y(y)}{(k+1)(1 + \sigma^2 \frac{\mathrm{d}^2}{\mathrm{d}y^2} \log f_Y(y))}|_{y=a}, k \ge 1. \tag{46}$$

Remark 5. While the formula for the coefficient in Theorem 4 is algebraically involved, in principle, it is not difficult to implement numerically.

[1]The identity in (36) is not new and was shown previously by Hatsell and Nolte in [8].

V. On the Distribution of the Conditional Expectation

Combing Theorem 4 with (5) and (6) we arrive at the expression for the distribution of $\hat{X}$. The following theorem summarizes this result.

Theorem 5. *Let $F_{\hat{X}}$ and $f_{\hat{X}}(x)$ be the cdf and pdf of the conditional mean estimator $\hat{X}(Y)$. Then, for a non-constant X and $\sigma > 0$, we have that*

$$F_{\hat{X}}(x) = F_Y(\hat{X}^{-1}(x)), \tag{47}$$

$$f_{\hat{X}}(x) = f_Y\left(\hat{X}^{-1}(x)\right)\left|\frac{\mathrm{d}}{\mathrm{d}x}\hat{X}^{-1}(x)\right|, \tag{48}$$

where

- *F_Y and f_Y are the cdf and the pdf of Y, respectively; and*
- *the inverse $\hat{X}^{-1}(x)$ is well-defined and given in Theorem 4. Furthermore, $\hat{X}^{-1}(x)$ can be expressed only in terms f_Y and derivatives of f_Y.*

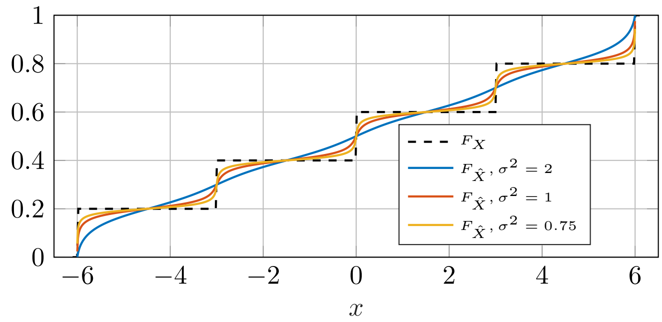
We now consider a few examples. To implement these examples, we truncate the power series in Theorem 4 such that the absolute error in the approximation is always below 10^{-4}.

- (Example with a Discrete Distribution) Let X be uniformly distributed on $\{-6, -3, 0, 3, 6\}$. Fig. 3a shows the cdf of $\hat{X}$ for several values of σ^2.
- (Mismatched Example 1) Let $\hat{X}(y; \sigma_m^2)$ denote the estimator that assumes that the noise variance is σ_m^2. Suppose, however, that $\hat{X}(y; \sigma_m^2)$ is used when the true noise level is $\sigma^2 \neq \sigma_m^2$. This scenario is known as mismatched estimation. Fig. 3b show the distribution of $\hat{X}(Y; \sigma_m^2)$ where $Y = X + N$ and $N \sim \mathcal{N}(0, \sigma^2)$. Moreover, X is uniformly distributed on $\{-6, -3, 0, 3, 6\}$.
- (Mismatched Example 2) Let $\hat{X}(y; Q_X)$ denote the estimator that assumes that the distribution of X is Q_X. Suppose, however, that $\hat{X}(y; Q_X)$ is used when the true the distribution is $P_X \neq Q_X$. In this scenario, we assume that the noise variance is used correctly. Fig. 3c show the distribution of $\hat{X}(Y; Q_X)$ where $Y = X + N$ and $X \sim P_X$. We assume that P_X is uniform over $\{-3, 0, 3\}$ while Q_X has the following assignment over the set $\{-3, 0, 3\}$: $Q_X(-3) = Q_X(3) = \frac{1-p}{2}$ and $Q_X(0) = p$.

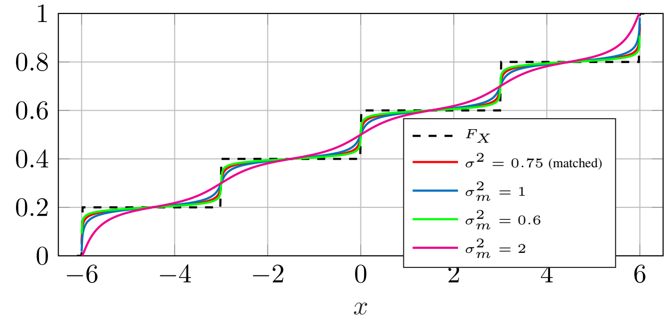
VI. Conclusion

This work has characterized the distribution of the conditional mean estimator of a random variable X based on a noisy Gaussian observation Y. This distribution has been shown to depend on the joint distribution P_{XY} only through the marginal of Y. Several new results have been shown along the way, such as a new identity between conditional expectation and conditional cumulants, and an inverse of the conditional expectation has been characterized in terms of a power series.

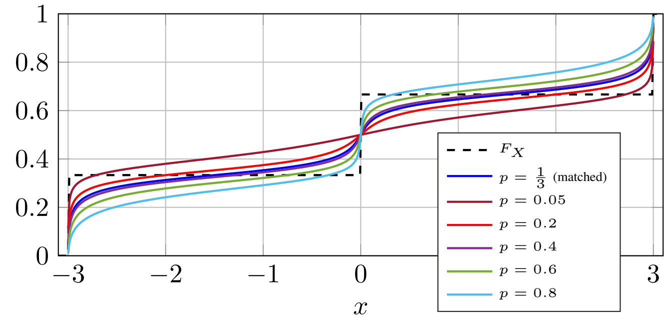
An interesting future direction would be to consider the distribution of the estimation error $X - \mathbb{E}[X|Y]$. Moreover, it would be interesting to see if the distribution of $\mathbb{E}[X|Y]$ can be used with the I-MMSE relationship [10] to characterize some features of the mutual information.



(a) Plots of $F_{\hat{X}}$ and F_X where X uniformly distributed over the set $\{-6, -3, 0, 3, 6\}$.



(b) Plots of the cdf of $\hat{X}(Y; \sigma_m^2)$ for several value of σ_m^2 where X is uniformly distributed over the set $\{-6, -3, 0, 3, 6\}$ and where $\sigma^2 = 0.75$.



(c) Plots of the cdf of $\hat{X}(Y; Q_X)$ where P_X is uniform over $\{-3, 0, 3, \}$ and where $Q_X(-3) = Q_X(3) = \frac{1-p}{2}$ and $Q_X(0) = p$ and where $\sigma^2 = 0.75$.

Fig. 3: Examples of the cdf of $\hat{X}$.

References

[1] A. Dytso, H. V. Poor, and S. Shamai (Shitz), "On the distribution of the conditional mean estimator in Gaussian noise," 2020. [Online]. Available: https://web.njit.edu/~ad94/ITW2021.pdf

[2] T. Han, *Information-Spectrum Methods in Information Theory*. Springer Science & Business Media, 2013, vol. 50.

[3] H. Robbins, "An empirical Bayes approach to statistics," in *Proc. Third Berkeley Symp. Math Statist. Probab.* Citeseer, 1956.

[4] R. Esposito, "On a relation between detection and estimation in decision theory," *Inf. Control*, vol. 12, no. 2, pp. 116–120, 1968.

[5] A. Dytso, H. V. Poor, and S. Shamai (Shitz), "A general derivative identity for the conditional mean estimator in Gaussian noise and some applications," in *Proc. IEEE Int. Symp. Inf. Theory*, Los Angeles, USA, June 2020, pp. 1183–1188.

[6] C. A. Charalambides, *Enumerative Combinatorics*. CRC Press, 2018.

[7] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, 2013.

[8] C. Hatsell and L. Nolte, "Some geometric properties of the likelihood ratio (corresp.)," *IEEE Trans. Inf. Theory*, vol. 17, no. 5, pp. 616–618, 1971.

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

[10] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.