Entropy **2012**, *14*, 966-977; doi:10.3390/e14050966

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Some Further Results on the Minimum Error Entropy Estimation

Badong Chen [1,2,*] and Jose C. Principe [2]

[1] Department of Precision Instruments and Mechanology, Tsinghua University, Beijing, 100084, China

[2] Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA; E-Mail: principe@cnel.ufl.edu

* Author to whom correspondence should be addressed; E-Mail: chenbd04@mails.tsinghua.edu.cn.

Received: 1 April 2012; in revised form: 2 May 2012 / Accepted: 10 May 2012 / Published: 21 May 2012

Abstract: The minimum error entropy (MEE) criterion has been receiving increasing attention due to its promising perspectives for applications in signal processing and machine learning. In the context of Bayesian estimation, the MEE criterion is concerned with the estimation of a certain random variable based on another random variable, so that the error's entropy is minimized. Several theoretical results on this topic have been reported. In this work, we present some further results on the MEE estimation. The contributions are twofold: (1) we extend a recent result on the minimum entropy of a mixture of unimodal and symmetric distributions to a more general case, and prove that if the conditional distributions are generalized uniformly dominated (GUD), the dominant alignment will be the MEE estimator; (2) we show by examples that the MEE estimator (not limited to singular cases) may be non-unique even if the error distribution is restricted to zero-mean (unbiased).

Keywords: entropy; estimation; minimum error entropy estimation

MSC Codes: 62B10

---

1. Introduction

A central concept in information theory is entropy, which is a mathematical measure of the uncertainty or the amount of missing information [1]. Entropy has been widely used in many areas, including physics, mathematics, communication, economics, signal processing, machine learning, *etc.* The maximum entropy principle is a powerful and widely accepted method for statistical inference or probabilistic reasoning with incomplete knowledge of probability distribution [2]. Another important entropy principle is the minimum entropy principle, which decreases the uncertainty associated with a system. In particular, the minimum error entropy (MEE) criterion can be applied in problems like estimation [3–5], identification [6,7], filtering [8–10], and system control [11,12]. In recent years, the MEE criterion, together with the nonparametric Renyi entropy estimator, has been successfully used in information theoretic learning (ITL) [13–15].

In the scenario of Bayesian estimation, the MEE criterion aims to minimize the entropy of the estimation error, and hence decrease the uncertainty in estimation. Given two random variables: $X \in \mathbb{R}^n$, an unknown parameter to be estimated, and $Y \in \mathbb{R}^m$, the observation (or measurement), the MEE estimation of X based on Y can be formulated as:

$$\begin{aligned} g^* &= \arg\min_{g \in \mathcal{G}} H\left(X - g(Y)\right) \\ &= \arg\min_{g \in \mathcal{G}} E\left[-\log p^g\left(X - g(Y)\right)\right] \\ &= \arg\min_{g \in \mathcal{G}} -\int_{\mathbb{R}^n} p^g(x) \log p^g(x) dx \end{aligned} \tag{1}$$

where $g(Y)$ denotes an estimator of X based on Y, g is a measurable function, $\mathcal{G}$ stands for the collection of all measurable functions of Y, $H\left(X - g(Y)\right)$ denotes the Shannon entropy of the estimation error $X - g(Y)$, and $p^g(x)$ denotes the probability density function (PDF) of the estimation error. Let $p(x|y)$ be the conditional PDF of X given Y. Then:

$$p^g(x) = \int_{\mathbb{R}^m} p(x + g(y) \mid y) dF(y) \tag{2}$$

where $F(y)$ denotes the distribution function of Y. From (2), one can see the error PDF $p^g(x)$ is actually a mixture of the shifted conditional PDF.

Different from conventional Bayesian risks, like mean square error (MSE) and risk-sensitive cost [16], the "loss function" in MEE is $-\log p^g(.)$, which is directly related to the error's PDF, transforming nonlinearly the error by its own PDF. Some theoretical aspects of MEE estimation have been studied in the literature. In an early work [3], Weidemann and Stear proved that minimizing the error entropy is equivalent to minimizing the mutual information between the error and the observation, and also proved that the reduced error entropy is upper-bounded by the amount of information obtained by the observation. In [17], Janzura *et al.* proved that, for the case of finite mixtures (Y is a discrete random variable with finite possible values), the MEE estimator equals the conditional median provided that the conditional PDFs are conditionally symmetric and unimodal (CSUM). Otahal [18] extended Janzura's results to finite-dimensional Euclidean space. In a recent paper, Chen and Geman [19] employed a "function rearrangement" to study the minimum entropy of a mixture of CSUM distributions where no restriction on Y was imposed. More recently, Chen *et al.* have

investigated the robustness, non-uniqueness (for singular cases), sufficient condition, and the necessary condition involved in the MEE estimation [20]. Chen *et al.* have also presented a new interpretation on the MSE criterion as a robust MEE criterion [21].

In this work, we continue the study on the MEE estimation, and obtain some further results. Our contributions are twofold. First, we extend the results of Chen and Geman to a more general case, and show that when the conditional PDFs are *generalized uniformly dominated* (GUD), the MEE estimator equals the *dominant alignment*. Second, we show by examples that, the unbiased MEE estimator (not limited to singular cases) may be non-unique, and there can even be infinitely many optimal solutions. The rest of the paper is organized as follows. In Section 2, we study the minimum entropy of a mixture of generalized uniformly dominated conditional distributions. In Section 3, we present two examples to show the non-uniqueness of the unbiased MEE estimation. Finally, we give our conclusions in Section 4.

2. MEE Estimator for Generalized Uniformly Dominated Conditional Distributions

Before presenting the main theorem of this section, we give the following definitions.

Definition 1: Let $\mathscr{F} = \left\{ f_t(x) : \mathbb{R}^n \to \mathbb{R}_+, t \in \mathrm{T} \right\}$ be a set of nonnegative, integrable functions, where T denotes an index set (possibly uncountable). Then $\mathscr{F}$ is said to be *uniformly dominated* in $x \in \mathbb{R}^n$ if and only if $\forall \varepsilon \in [0, \infty)$, there exists a measurable set $D_\varepsilon \subset \mathbb{R}^n$, satisfying $\lambda(D_\varepsilon) = \varepsilon$ and:

$$\int_{D_\varepsilon} f_t(x)\,dx = \sup_{A:\lambda(A)=\varepsilon} \int_A f_t(x)\,dx, \ \forall t \in \mathrm{T} \tag{3}$$

where λ is Lebesgue measure. The set D_ε is called the ε*-volume dominant support* of $\mathscr{F}$.

Definition 2: *The* nonnegative, integrable function set $\mathscr{F}$ is said to be *generalized uniformly dominated* (GUD) in $x \in \mathbb{R}^n$ if and only if there exists a function $\pi : \mathrm{T} \to \mathbb{R}^n$, such that $\mathscr{F}_\pi$ is uniformly dominated, where:

$$\mathscr{F}_\pi = \left\{ f_t^\pi : f_t^\pi(x) = f_t(x + \pi(t)), t \in \mathrm{T} \right\} \tag{4}$$

The function $\pi(t)$ is called the *dominant alignment* of $\mathscr{F}$.

Remark 1: The dominant alignment is, obviously, non-unique. If $\pi(t)$ is a dominant alignment of $\mathscr{F}$, then $\forall \boldsymbol{c} \in \mathbb{R}^n$, $\pi(t) + \boldsymbol{c}$ will also be a dominant alignment of $\mathscr{F}$.

When regarding y as an *index* parameter, the conditional PDF $p(x|y)$ will represent a set of nonnegative and integrable functions, that is:

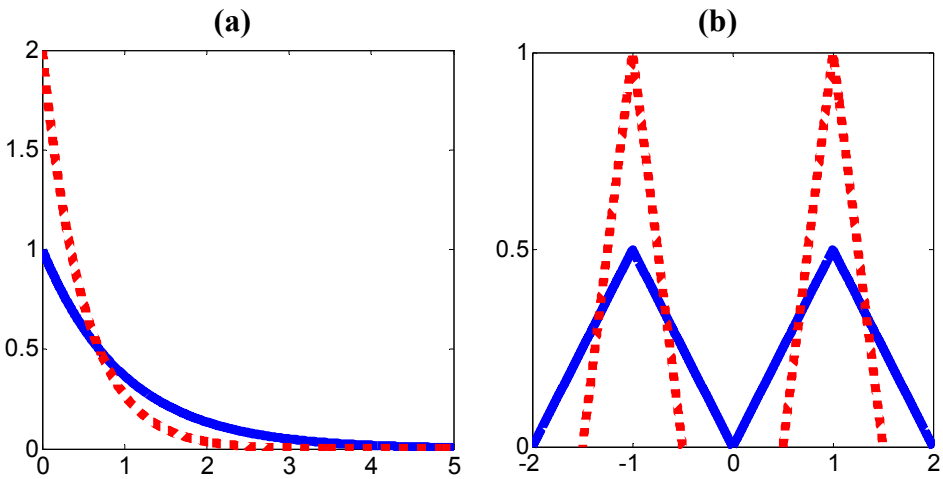
$$\mathscr{P} = \left\{ p(x|y) : \mathbb{R}^n \to \mathbb{R}_+, y \in \mathbb{R}^m \right\} \tag{5}$$

If the above function set is (generalized) uniformly dominated in $x \in \mathbb{R}^n$, then we say that the conditional PDF $p(x|y)$ is (generalized) uniformly dominated in $x \in \mathbb{R}^n$.

Remark 2: The GUD is much more general than CSUM. Actually, if the conditional PDF is CSUM, it must also be GUD (with the conditional mean as the dominant alignment), but not *vice versa*. In

Figure 1 we show two examples where two PDFs (solid and dotted lines) are uniformly dominated but not CSUM.

Figure 1. Uniformly dominated PDFs: (**a**) non-symmetric, (**b**) non-unimodal.



Theorem 1: Assume the conditional PDF $p(x|y)$ is generalized uniformly dominated in $x \in \mathbb{R}^n$, with dominant alignment $\pi(y)$. If $H(X-\pi(Y))$ exists (here "exists" means "exists in the extended sense" as defined in [19]) then $H(X-\pi(Y)) \le H(X-g(Y))$ for all $g:\mathbb{R}^m \to \mathbb{R}^n$ for which $H(X-g(Y))$ also exists.

Proof of Theorem 1: The proof presented *below* is similar to that of the *Theorem 1* in [19], except that the discretization procedure is avoided. In the following, we give a brief sketch of the proof, and consider only the case $n=1$ (the proof can be easily extended to $n>1$). First, one needs to prove the following proposition.

Proposition 1: Assume that the function $f(x|y)$, $x \in \mathbb{R}$, $y \in \mathbb{R}^m$ (not necessarily a conditional PDF) satisfies the following conditions:

(1) non-negative, continuous, and integrable in $x \in \mathbb{R}$ for each $y \in \mathbb{R}^m$;
(2) generalized uniformly dominated in x, with dominant alignment $\pi(y)$;
(3) uniformly bounded in (x, y).

Then *for* any $g:\mathbb{R}^m \to \mathbb{R}$, we have:

$$H\left(f^{\pi}\right) \le H\left(f^{g}\right) \tag{6}$$

where $H(f) = -\int_{\mathbb{R}} f(x)\log f(x)dx$ (here we extend the entropy definition to nonnegative L^1 functions), and:

$$\begin{cases} f^{\pi}(x) = \int_{\mathbb{R}^m} f(x+\pi(y)\,|\,y)dF(y) \\ f^{g}(x) = \int_{\mathbb{R}^m} f(x+g(y)\,|\,y)dF(y) \end{cases} \tag{7}$$

Remark 3: It is easy to verify that $\int f^{\pi}dx = \int f^{g}dx \le \sup_{(x,y)} f\left(x|y\right) < \infty$ (not necessarily $\int f^{\pi}dx = 1$).

Proof of Proposition 1: The above proposition can be readily proved using the following two lemmas.

Lemma 1 [19]: Assume the nonnegative function $h:\mathbb{R} \to [0,\infty)$ is bounded, continuous, and integrable, and define the function $O_h(z)$ by (λ is Lebesgue measure):

$$O_h(z) = \lambda\{x : h(x) \ge z\} \tag{8}$$

Then the following results hold:

(a) Define $m^h(x) = \sup\{z : O_h(z) \ge x\}$, $x \in (0,\infty)$, and $m^h(0) = \sup_x h(x)$. Then $m^h(x)$ is continuous and non-increasing on $[0,\infty)$, and $m^h(x) \to 0$ as $x \to \infty$.

(b) For any function $G:[0,\infty) \to \mathbb{R}$ with $\int_{\mathbb{R}} |G(h(x))|dx < \infty$

$$\int_{\mathbb{R}} G(h(x))dx = \int_0^{\infty} G(m^h(x))dx \tag{9}$$

(c) For any $x_0 \in [0,\infty)$

$$\int_0^{x_0} m^h(x)dx = \sup_{A:\lambda(A)=x_0} \int_A h(x)dx \tag{10}$$

Proof of Lemma 1: See [19].

Remark 4: Denote $m^{\pi} = m^{f^{\pi}}$, and $m^{g} = m^{f^{g}}$. Then by *Lemma 1*, we have $H(m^{\pi}) = H(f^{\pi})$ and $H(m^{g}) = H(f^{g})$ (let $G(x) = -x\log x$). Therefore, to prove *Proposition 1*, it suffices to prove:

$$H(m^{\pi}) \le H(m^{g}) \tag{11}$$

Lemma 2: Functions m^{π} and m^{g} satisfy:

(a)

$$\int_0^{\infty} m^{\pi}(x)dx = \int_0^{\infty} m^{g}(x)dx < \infty \tag{12}$$

(b)

$$\int_0^{\varepsilon} m^{g}(x)dx \le \int_0^{\varepsilon} m^{\pi}(x)dx, \quad \forall \varepsilon \in [0,\infty) \tag{13}$$

Proof of Lemma 2: (a) *comes* directly from the fact $\int f^{\pi}dx = \int f^{g}dx < \infty$. We only need to prove (b). We have:

$$\begin{aligned}
&\sup_{A:\lambda(A)=\varepsilon} \int_A f^g(x)dx \\
&= \sup_{A:\lambda(A)=\varepsilon} \int_A \int_{\square^m} f(x+g(y)\,|\,y)dF(y)dx \\
&= \sup_{A:\lambda(A)=\varepsilon} \int_{\square^m} \int_A f(x+g(y)\,|\,y)dxdF(y) \\
&\le \int_{\square^m} \left\{ \sup_{A:\lambda(A)=\varepsilon} \int_A f(x+g(y)\,|\,y)dx \right\} dF(y) \\
&= \int_{\square^m} \left\{ \int_{D_\varepsilon} f(x+\pi(y)\,|\,y)dx \right\} dF(y) \\
&= \int_{D_\varepsilon} \int_{\square^m} f(x+\pi(y)\,|\,y)dF(y)dx \\
&= \sup_{A:\lambda(A)=\varepsilon} \int_A \int_{\square^m} f(x+\pi(y)\,|\,y)dF(y)dx \\
&= \sup_{A:\lambda(A)=\varepsilon} \int_A f^{\pi}(x)dx
\end{aligned} \tag{14}$$

where D_ε is the ε-volume dominant support of $f(x+\pi(y)\,|\,y)$. By *Lemma 1* (c):

$$\begin{cases}
\int_0^{\varepsilon} m^g(x)dx = \sup_{A:\lambda(A)=\varepsilon} \int_A f^g(x)dx \\
\int_0^{\varepsilon} m^{\pi}(x)dx = \sup_{A:\lambda(A)=\varepsilon} \int_A f^{\pi}(x)dx
\end{cases} \tag{15}$$

Thus $\forall \varepsilon \in [0,\infty)$, we have $\int_0^{\varepsilon} m^g(x)dx \le \int_0^{\varepsilon} m^{\pi}(x)dx$:

Q.E.D (*Lemma 2*)

We are now in position to prove (11):

$$\begin{aligned}
H\left(m^g\right) &= -\int_0^{\infty} m^g(x)\log m^g(x)dx \\
&= \int_0^{\infty} m^g(x)\int_0^{-\log m^g(x)} dydx \\
&= \int_0^{\infty}\left(\int_0^{\infty} m^g(x) I\left[y \le -\log m^g(x)\right]dy\right)dx \\
&= \int_0^{\infty}\left(\int_{\inf\{x:-\log m^g(x)\ge y\}}^{\infty} m^g(x)dx\right)dy \\
&\overset{Lemma\,2}{\ge} \int_0^{\infty}\left(\int_{\inf\{x:-\log m^g(x)\ge y\}}^{\infty} m^{\pi}(x)dx\right)dy \\
&= -\int_0^{\infty} m^{\pi}(x)\log m^g(x)dx \\
&\overset{(a)}{\ge} -\int_0^{\infty} m^{\pi}(x)\log m^{\pi}(x)dx \\
&= H(m^{\pi})
\end{aligned} \tag{16}$$

where $I\left[.\right]$ denotes the indicator function, and (a) follows from $\forall x > 0$, $-\log x \ge 1-x$, that is:

$$
\begin{aligned}
&-\int_0^\infty m^\pi(x)\log m^g(x)dx+\int_0^\infty m^\pi(x)\log m^\pi(x)dx \\
&=-\int_0^\infty m^\pi(x)\log\left(\frac{m^g(x)}{m^\pi(x)}\right)dx \\
&\geq \int_0^\infty m^\pi(x)\left(1-\frac{m^g(x)}{m^\pi(x)}\right)dx \\
&=\int_0^\infty m^\pi(x)dx-\int_0^\infty m^g(x)dx \overset{Lemma2}{=} 0
\end{aligned}
\tag{17}
$$

In the above proof, we adopt the convention $0\cdot\infty=0$.

Q.E.D (*Proposition 1*)

Now the proof of *Proposition 1* has been completed. To finish the proof of *Theorem 1*, we have to remove the conditions of continuity and uniform boundedness imposed in *Proposition 1*. This can be easily accomplished by approximating $p(x\,|\,y)$ by a sequence of functions $\{f_n(x\,|\,y)\}$, $n=1,2,\cdots$, which satisfy these conditions. The remaining proof is omitted here, since it is exactly the same as the last part of the proof for *Theorem 1* in [19].

Q.E.D. (*Theorem 1*)

Example 1: *Consider* an additive noise model:

$$X=\varphi(Y)+\xi \tag{18}$$

where ξ is an *additive* noise that is independent of Y. In this case, we have $p(x\,|\,y)=p_\xi(x-\varphi(y))$, where $p_\xi(.)$ denotes the noise PDF. It is clear that $p(x\,|\,y)$ is generalized uniformly dominated, with dominant alignment $\pi(y)=\varphi(y)$. According to *Theorem 1*, we have $H(X-\varphi(Y))\leq H(X-g(Y))$. In fact, this result can also be proved by:

$$
\begin{aligned}
H(X-\varphi(Y)) &= H(\xi) \\
&\overset{(b)}{\leq} H(\xi+(\varphi(Y)-g(Y))) \\
&= H(X-g(Y))
\end{aligned}
\tag{19}
$$

where (b) comes from the fact that ξ and $(\varphi(Y)-g(Y))$ are independent (For independent random variables X and Y, the inequality $H(X+Y)\geq\max\{H(X),H(Y)\}$ holds). In this example, the conditional PDF $p(x\,|\,y)$ is, obviously, not necessarily CSUM.

Example 2: *Suppose* the joint PDF of random variables X, Y ($X,Y\in\mathbb{R}$) is:

$$p(x,y)=\exp(1-x\exp(y)) \tag{20}$$

where $y\geq 0$, $x\exp(y)\geq 1$. Then the conditional PDF $p(x|y)$ will be:

$$
\begin{aligned}
p(x|y) &= \frac{p(x,y)}{p(y)} \\
&= \frac{\exp(1-x\exp(y))}{\int_{\exp(-y)}^{\infty} \exp(1-x\exp(y))\,dx} \\
&= \exp(1-x\exp(y)+y)
\end{aligned}
\tag{21}
$$

One can easily verify that the above conditional PDF is non-symmetric but generalized uniformly dominated, with dominant alignment $\pi(y)=\exp(-y)$ (the ε-volume dominant support of $p(x+\pi(y)\,|\,y)$ is $D_\varepsilon=[0,\varepsilon]$). By *Theorem 1*, the function $\exp(-y)$ is the minimizer of error entropy.

3. Non-Uniqueness of Unbiased MEE Estimation

Because entropy is shift-invariant, the MEE estimator is obviously non-unique. In practical applications, in order to yield a unique solution, or to meet the desire for small error values, the MEE estimator is usually restricted to be unbiased, that is, the estimation error is restricted to be zero-mean [15]. The question of interest in this paper is whether the unbiased MEE estimator is unique. In [20], it has been shown that, for the singular case (in which the error entropy approaches minus infinity), the unbiased MEE estimation may yield non-unique (even infinitely many) solutions. In the following, we present two examples to show that this result still holds even for nonsingular case.

Example 3: Let the joint PDF of X and Y ($X,Y\in\mathbb{R}$) be a mixed-Gaussian density [20]:

$$
p(x,y)=\frac{1}{4\pi\sqrt{1-\rho^2}}\left\{\exp\left\{\frac{-\left(y^2-2\rho y(x-\mu)+(x-\mu)^2\right)}{2\left(1-\rho^2\right)}\right\}+\exp\left\{\frac{-\left(y^2+2\rho y(x+\mu)+(x+\mu)^2\right)}{2\left(1-\rho^2\right)}\right\}\right\}
\tag{22}
$$

where $\mu>0$, $0\le|\rho|<1$. The conditional PDF of X given Y will be:

$$
p(x|y)=\frac{1}{2\sqrt{2\pi\left(1-\rho^2\right)}}\left\{\exp\left(-\frac{(x-\mu-\rho y)^2}{2\left(1-\rho^2\right)}\right)+\exp\left(-\frac{(x+\mu+\rho y)^2}{2\left(1-\rho^2\right)}\right)\right\}
\tag{23}
$$

$\forall y\in\mathbb{R}$, $p(x\,|\,y)$ is symmetric around zero (but not unimodal in *x*). It can be shown that for some values of μ, ρ, the MEE estimator of X based on Y does not equal zero (see [20], *Example 3*). In these cases, the MEE estimator will be non-unique, even if the error's PDF is restricted to zero-mean (unbiased) distribution. This can be proved as follows:

Let g^* be an unbiased MEE estimator of X based on Y. Then $-g^*$ will also be an unbiased MEE estimator, because:

$$
\begin{aligned}
p^{g^*}(x) &= \int_{\mathbb{R}} p(x+g^*(y)\,|\,y)dF(y) \\
&\overset{(c)}{=} \int_{\mathbb{R}} p(-x-g^*(y)\,|\,y)dF(y) \\
&= p^{-g^*}(-x)
\end{aligned}
\tag{24}
$$

where (c) comes from the fact that $p(x\,|\,y)$ is symmetric around zero, and further:

$$\begin{aligned} H\left(p^{g^*}\right) &= -\int_{-\infty}^{\infty} p^{g^*}(x)\log p^{g^*}(x)dx \\ &= -\int_{-\infty}^{\infty} p^{-g^*}(-x)\log p^{-g^*}(-x)dx \\ &= -\int_{-\infty}^{\infty} p^{-g^*}(x)\log p^{-g^*}(x)dx \\ &= H\left(p^{-g^*}\right) \end{aligned} \tag{25}$$

If the unbiased MEE estimator is unique, then we have $g^* = -g^* \Rightarrow g^* = 0$, which contradicts the fact that g^* does not equal zero. Therefore, the unbiased MEE estimator must be non-unique. Obviously, the above result can be extended to more general cases. In fact, we have the following proposition.

Proposition2: The unbiased MEE estimator will be non-unique if the conditional PDF $p(x|y)$ satisfies:

(1) Symmetric in $x \in \mathbb{R}^n$ around the conditional mean $\mu(y) = E\left[X|Y=y\right]$ for each $y \in \mathbb{R}^m$;

(2) There exists a function $\zeta : \mathbb{R}^m \to \mathbb{R}^n$ such that $\zeta(Y) \neq \mu(Y)$, $H\left(X-\zeta(Y)\right) \leq H\left(X-\mu(Y)\right)$.

Proof: Similar to the proof presented above (Omitted).

In the next example, we show that, for some particular situations, there can be even infinitely many unbiased MEE estimators.

Example 4: *Suppose* Y is a discrete random variable with Bernoulli distribution:

$$\Pr(Y=0) = \Pr(Y=1) = 0.5 \tag{26}$$

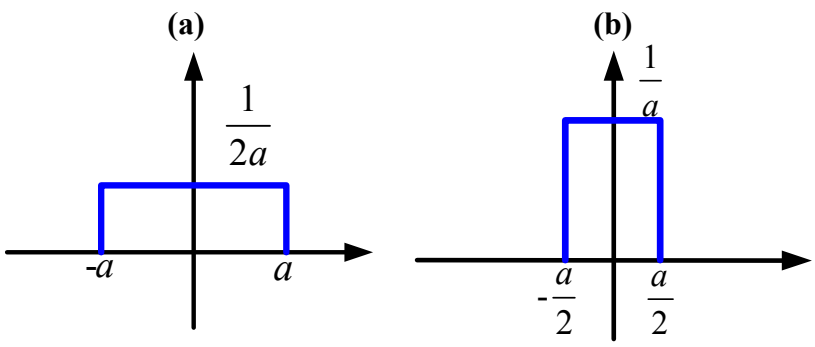
The conditional PDF of X given Y is (see Figure 2):

$$p\left(x|Y=0\right) = \begin{cases} \dfrac{1}{2a} & if \ |x| \leq a \\ 0 & other \end{cases}$$

$$p\left(x|Y=1\right) = \begin{cases} \dfrac{1}{a} & if \ |x| \leq \dfrac{1}{2}a \\ 0 & other \end{cases}$$

where $a > 0$. Note that the above conditional PDF is uniformly dominated in x.

Figure 2. Conditional PDF of X given Y: (**a**) $p(x|Y=0)$, (**b**) $p(x|Y=1)$.



Given an estimator $\hat{X} = g(Y)$, the error's PDF will be:

$$p^g(x) = \frac{1}{2}\left\{p\left(x + g(0)\middle|Y = 0\right) + p\left(x + g(1)\middle|Y = 1\right)\right\} \tag{27}$$

Let g be an unbiased estimator, then $\int_{-\infty}^{\infty} xp^g(x)dx = 0$, and hence $g(0) = -g(1)$. In the following, we assume $g(0) \geq 0$ (due to symmetry, one can obtain similar results for $g(0) < 0$), and consider three cases:

Case 1: $0 \leq g(0) \leq \frac{1}{4}a$. In this case, the error PDF is:

$$p^g(x) = \begin{cases} \frac{1}{4a}, & -a - g(0) \leq x < -\frac{a}{2} + g(0) \\ \frac{3}{4a}, & -\frac{a}{2} + g(0) \leq x \leq \frac{a}{2} + g(0) \\ \frac{1}{4a}, & \frac{a}{2} + g(0) < x \leq a - g(0) \end{cases} \tag{28}$$

Then the error entropy can be calculated as:

$$H\left(p^g(x)\right) = -\frac{3}{4}\log(3) + \log(4a) \tag{29}$$

Case 2: $\frac{1}{4}a < g(0) \leq \frac{3}{4}a$. In this case, we have:

$$p^g(x) = \begin{cases} \frac{1}{4a}, & -a - g(0) \leq x < -\frac{a}{2} + g(0) \\ \frac{3}{4a}, & -\frac{a}{2} + g(0) \leq x \leq a - g(0) \\ \frac{1}{2a}, & a - g(0) < x \leq \frac{a}{2} + g(0) \end{cases} \tag{30}$$

And hence:

$$H\left(p^g(x)\right) = \left(-\frac{9}{8} + \frac{3g(0)}{2a}\right)\log(3) + \left(\frac{1}{4} - \frac{g(0)}{a}\right)\log(2) + \log(4a) \tag{31}$$

Case 3: $g(0) > \frac{3}{4}a$. In this case:

$$p^g(x) = \begin{cases} \frac{1}{4a}, & -a - g(0) \leq x \leq a - g(0) \\ \frac{1}{2a}, & -\frac{a}{2} + g(0) \leq x \leq \frac{a}{2} + g(0) \end{cases} \tag{32}$$

Thus:

$$H\left(p^g(x)\right) = -\frac{1}{2}\log(2) + \log(4a) \tag{33}$$

One can easily verify that the error entropy achieves its minimum value when $0 \le g(0) \le a/4$ (the first case). There are, therefore, infinitely many unbiased estimators that minimize the error entropy.

4. Conclusion

Two issues involved in the minimum error entropy (MEE) estimation have been studied in this work. The first issue is about which estimator minimizes the error entropy. In general there is no explicit expression for the MEE estimator unless some constraints on the conditional distribution are imposed. In the past, several researchers have shown that, if the conditional density is conditionally symmetric and unimodal (CSUM), then the conditional mean (or median) will be the MEE estimator. We extend these results to a more general case, and show that if the conditional densities are generalized uniformly dominated (GUD), then the dominant alignment will minimize the error entropy. The second issue is about the non-uniqueness of the unbiased MEE estimation. It has been shown in a recent paper that for the singular case (in which the error entropy approaches minus infinity), the unbiased MEE estimation may yield non-unique (even infinitely many) solutions. In this work, we show by examples that this result still holds even for nonsingular case.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 60904054), NSF grant ECCS 0856441, and ONR N00014-10-1-0375.

References

1. Cover, T.M.; Thomas, J.A. *Element of Information Theory*; Wiley & Son, Inc.: New York, NY, USA, 1991.
2. Kapur, J.N.; Kesavan, H.K. *Entropy Optimization Principles with Applications*; Academic Press, Inc.: Boston, MA, USA, 1992.
3. Weidemann, H.L.; Stear, E.B. Entropy analysis of estimating systems. *IEEE Trans. Inform. Theor.* **1970**, *16*, 264–270.
4. Tomita, Y.; Ohmatsu, S.; Soeda, T. An application of the information theory to estimation problems. *Inf. Control* **1976**, *32*, 101–111.
5. Wolsztynski, E.; Thierry, E.; Pronzato, L. Minimum-entropy estimation in semiparametric models. *Signal Process.* **2005**, *85*, 937–949.
6. Guo, L.Z.; Billings, S.A.; Zhu, D.Q. An extended orthogonal forward regression algorithm for system identification using entropy. *Int. J. Control* **2008**, *81*, 690–699.
7. Chen, B.; Zhu, Y.; Hu, J.; Principe, J.C. Δ-entropy: Definition, properties and applications in system identification with quantized data. *Inf. Sci.* **2011**, *181*, 1384–1402.
8. Kalata, P.; Priemer, R. Linear prediction, filtering and smoothing: an information theoretic approach. *Inf. Sci.* **1979**, *17*, 1–14.
9. Feng, X.; Loparo, K.A.; Fang, Y. Optimal state estimation for stochastic systems: An information theoretic approach. *IEEE Trans. Automat. Contr.* **1997**, *42*, 771–785.

10. Guo, L.; Wang, H. Minimum entropy filtering for multivariate stochastic systems with non-Gaussian Noises. *IEEE Trans. Autom. Control* **2006**, *51*, 695–700.
11. Wang, H. Minimum entropy control of non-Gaussian dynamic stochastic systems. *IEEE Trans. Autom. Control* **2002**, *47*, 398–403.
12. Yue, H.; Wang, H. Minimum entropy control of closed-loop tracking errors for dynamic stochastic systems. *IEEE Trans. Autom. Control* **2003**, *48*, 118–122.
13. Erdogmus, D.; Principe, J.C. An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems. *IEEE Trans. Signal Process.* **2002**, *50*, 1780–1786.
14. Erdogmus, D.; Principe, J.C. From linear adaptive filtering to nonlinear information processing—The design and analysis of information processing systems. *IEEE Signal Process. Mag.* **2006**, *23*, 14–33.
15. Principe, J.C. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*; Springer: New York, NY, USA, 2010.
16. Boel, R.K.; James M.R.; Petersen I.R. Robustness and risk-sensitive filtering. *IEEE Trans. Autom. Control* **2002**, *47*, 451–461.
17. Janzura, M.; Koski, T.; Otahal, A. Minimum entropy of error principle in estimation. *Inf. Sci.* **1994**, *79*, 123–144.
18. Otahal, A. Minimum entropy of error estimate for multi-dimensional parameter and finite-state-space observations. *Kybernetika* **1995**, *31*, 331–335.
19. Chen, T.-L.; Geman, S. On the minimum entropy of a mixture of unimodal and symmetric distributions. *IEEE Trans. Inf. Theory* **2008**, *54*, 3166–3174.
20. Chen, B.; Zhu, Y.; Hu, J.; Zhang, M. On optimal estimations with minimum error entropy criterion. *J. Frankl. Inst.-Eng. Appl. Math.* **2010**, *347*, 545–558.
21. Chen, B.; Zhu, Y.; Hu, J.; Zhang, M. A new interpretation on the MMSE as a robust MEE criterion. *Signal Process.* **2010**, *90*, 3313–3316.