EXPONENTIALLY EMBEDDED FAMILIES FOR MULTIMODAL SENSOR PROCESSING

Steven Kay, Quan Ding

Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island
Kelley Hall, 4 East Alumni Avenue, Kingston, RI 02881, Email: kay@ele.uri.edu, dingqqq@ele.uri.edu

ABSTRACT

The exponential embedding of two or more probability density functions (PDFs) is proposed for multimodal sensor processing. It approximates the unknown PDF by exponentially embedding the known PDFs. Such embedding is of a exponential family indexed by some parameters, and hence inherits many nice properties of the exponential family. It is shown that the approximated PDF is asymptotically the one that is the closest to the unknown PDF in Kullback-Leibler (KL) divergence. Applied to hypothesis testing, this approach shows improved performance compared to existing methods for cases of practical importance where the sensor outputs are not independent.

Index Terms— Sensor fusion, hypothesis testing, exponential embedding, exponential family, Kullback-Leibler divergence

1. INTRODUCTION

Distributed detection systems have many applications such as radar and sonar, medical diagnosis, weather prediction, and financial analysis. To obtain optimal performance, we require the joint PDF of the sensor outputs, which is not always available. One common approach [1], [2] is to assume that the PDFs of the sensor outputs are independent, and hence the joint PDF is the product of the marginal PDFs. However, this assumption may not be satisfied since the sensor measurements could be correlated due to the common source and the relative sensor locations. The correlation is noticed in [3], [4], where a copula based framework is proposed to estimate the joint PDF from the marginal PDFs. In this work, we approximate the joint PDF by exponentially embedded families (EEFs) in the sense that it asymptotically minimizes the KL divergence of the true PDF and the estimated one. For two PDFs p_1 and p_0, the KL divergence is defined as

$$D\left(p_1 \left\| p_0\right.\right) = \int p_1(\mathbf{x}) \ln \frac{p_1(\mathbf{x})}{p_0(\mathbf{x})} d\mathbf{x}$$

It is always nonnegative and equals zero if and only if $p_1 = p_0$ almost everywhere. The KL divergence is a measure of the asymptotic performance of binary hypothesis testing by Stein's lemma [5].

This work was supported by the Air Force Office of Scientific Research through the Air Force Research Laboratory Sensors Directorate under contract number FA8650-08-D-1303 0006.

The term "exponentially embedded familie" follows that in [6], where it is used for model order estimation. The embedded PDF is of an exponential family indexed by one or more parameters, and so has many nice properties of that family. In a differential geometry point of view, the EEF forms a manifold in log-PDF space. In one-dimensional case, the EEF is the PDF that minimizes $D\left(p \left\| p_0\right.\right)$ with the constraint that $D\left(p \left\| p_0\right.\right) - D\left(p \left\| p_1\right.\right) = \theta$ [5], [7]. Here we focus on the problem of binary hypothesis testing. We assume the presence of two sensors in this paper. Similar results are obtained for multiple hypothesis testing and multiple sensors.

The paper is organized as follows. Section 2 defines the EEF and discusses its properties. Followed in Section 3 is its application for hypothesis testing. An example is given in section 4. In Section 5, we show the simulation results by comparing the ROC curves of different approaches. Conclusion is drawn finally in Section 6.

2. EEF AND ITS PROPERTIES

Assume that a source produces the underlying samples $\mathbf{x}$ which are unobservable, and we have two sensors whose outputs are the statistics $\mathbf{t}_1(\mathbf{x})$ and $\mathbf{t}_2(\mathbf{x})$ of $\mathbf{x}$. Consider the binary hypothesis testing problem where we know the reference PDF $p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)$, but not $p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_1)$. So we can find the joint PDF $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathcal{H}_0)$, but do not know $p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathcal{H}_1)$. We assume that the marginal PDFs $p_{\mathbf{T}_1}(\mathbf{t}_1;\mathcal{H}_1)$ and $p_{\mathbf{T}_2}(\mathbf{t}_2;\mathcal{H}_1)$ are known. So the problem is to test between $\mathcal{H}_0$ and $\mathcal{H}_1$ where we know the joint PDF under $\mathcal{H}_0$ and the marginal PDFs under $\mathcal{H}_1$. The EEF is defined as

$$p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta}) = \frac{\left(\frac{p_{\mathbf{T}_1}(\mathbf{t}_1(\mathbf{x});\mathcal{H}_1)}{p_{\mathbf{T}_1}(\mathbf{t}_1(\mathbf{x});\mathcal{H}_0)}\right)^{\eta_1} \left(\frac{p_{\mathbf{T}_2}(\mathbf{t}_2(\mathbf{x});\mathcal{H}_1)}{p_{\mathbf{T}_2}(\mathbf{t}_2(\mathbf{x});\mathcal{H}_0)}\right)^{\eta_2} p_{\mathbf{X}}\left(\mathbf{x};\mathcal{H}_0\right)}{\int \left(\frac{p_{\mathbf{T}_1}(\mathbf{t}_1(\mathbf{x});\mathcal{H}_1)}{p_{\mathbf{T}_1}(\mathbf{t}_1(\mathbf{x});\mathcal{H}_0)}\right)^{\eta_1} \left(\frac{p_{\mathbf{T}_2}(\mathbf{t}_2(\mathbf{x});\mathcal{H}_1)}{p_{\mathbf{T}_2}(\mathbf{t}_2(\mathbf{x});\mathcal{H}_0)}\right)^{\eta_2} p_{\mathbf{X}}\left(\mathbf{x};\mathcal{H}_0\right) d\mathbf{x}} \quad (1)$$

where $\boldsymbol{\eta} = [\eta_1,\ \eta_2]^T$ are the embedding parameters with the constraints

$$\boldsymbol{\eta} \in \{\boldsymbol{\eta} : \eta_1, \eta_2 \geq 0,\ \eta_1 + \eta_2 \leq 1\} = S \quad (2)$$

Notice that $p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta})$ does not require the knowledge of $p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_1)$. So in practice, we just need to estimate $p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)$ and only the PDFs of $\mathbf{T}_1$ and $\mathbf{T}_2$ under $\mathcal{H}_1$ from training data (see also [8]). The reason why we have the constraints in (2) will be explained later. The next theorem is an extension of Kullback's results [5], [7]..

Theorem. *The PDF of* $\mathbf{x}$ *as in (1) is the one that minimizes* $D\left(p_{\mathbf{X}}(\mathbf{x})\,\|p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)\right)$ *subject to the constraints that*

$$D\left(p_{\mathbf{T}_i}(\mathbf{t}_i)\,\|p_{\mathbf{T}_i}(\mathbf{t}_i;\mathcal{H}_0)\right)-D\left(p_{\mathbf{T}_i}(\mathbf{t}_i)\,\|p_{\mathbf{T}_i}(\mathbf{t}_i;\mathcal{H}_1)\right)=\theta_i$$

for $i=1,2$*, where* $p_{\mathbf{T}_1}(\mathbf{t}_1)$ *and* $p_{\mathbf{T}_2}(\mathbf{t}_2)$ *are the PDFs of* $\mathbf{T}_1$ *and* $\mathbf{T}_2$ *corresponding to* $p_{\mathbf{X}}(\mathbf{x})$*.*

Proof. Since

$$\begin{aligned}&D\left(p_{\mathbf{T}_i}(\mathbf{t}_i)\,\|p_{\mathbf{T}_i}(\mathbf{t}_i;\mathcal{H}_0)\right)-D\left(p_{\mathbf{T}_i}(\mathbf{t}_i)\,\|p_{\mathbf{T}_i}(\mathbf{t}_i;\mathcal{H}_1)\right)\\&=\int p_{\mathbf{X}}(\mathbf{x})\ln\frac{p_{\mathbf{T}_i}\left(\mathbf{t}_i(\mathbf{x});\mathcal{H}_1\right)}{p_{\mathbf{T}_i}\left(\mathbf{t}_i(\mathbf{x});\mathcal{H}_0\right)}d\mathbf{x}\ \text{ for } i=1,2\end{aligned}$$

using Lagrange multipliers for the minimization gives

$$\begin{aligned}&J\left(p_{\mathbf{X}}(\mathbf{x})\right)=\int p_{\mathbf{X}}(\mathbf{x})\ln\frac{p_{\mathbf{X}}(\mathbf{x})}{p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)}d\mathbf{x}\\&+\lambda_1\int p_{\mathbf{X}}(\mathbf{x})\ln\frac{p_{\mathbf{T}_1}\left(\mathbf{t}_1(\mathbf{x});\mathcal{H}_1\right)}{p_{\mathbf{T}_1}\left(\mathbf{t}_1(\mathbf{x});\mathcal{H}_0\right)}d\mathbf{x}\\&+\lambda_2\int p_{\mathbf{X}}(\mathbf{x})\ln\frac{p_{\mathbf{T}_2}\left(\mathbf{t}_2(\mathbf{x});\mathcal{H}_1\right)}{p_{\mathbf{T}_2}\left(\mathbf{t}_2(\mathbf{x});\mathcal{H}_0\right)}d\mathbf{x}+\lambda_3\int p_{\mathbf{X}}(\mathbf{x})d\mathbf{x}\end{aligned}$$

Differentiating with respect to $p_{\mathbf{X}}(\mathbf{x})$ and setting to 0, we have

$$\begin{aligned}&\ln\frac{p_{\mathbf{X}}(\mathbf{x})}{p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)}+1+\lambda_1\ln\frac{p_{\mathbf{T}_1}\left(\mathbf{t}_1(\mathbf{x});\mathcal{H}_1\right)}{p_{\mathbf{T}_1}\left(\mathbf{t}_1(\mathbf{x});\mathcal{H}_0\right)}\\&+\lambda_2\ln\frac{p_{\mathbf{T}_2}\left(\mathbf{t}_2(\mathbf{x});\mathcal{H}_1\right)}{p_{\mathbf{T}_2}\left(\mathbf{t}_2(\mathbf{x});\mathcal{H}_0\right)}+\lambda_3=0\end{aligned}$$

Solving this equation and letting $\eta_1=-\lambda_1$ and $\eta_2=-\lambda_2$, the $p_{\mathbf{X}}(\mathbf{x})$ that minimizes $D\left(p_{\mathbf{X}}(\mathbf{x})\,\|p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)\right)$ is of the form as in (1) where η_1 and η_2 are chosen to meet the constraints. □

By letting

$$\begin{aligned}K(\boldsymbol{\eta})=\ln\int&\left(\frac{p_{\mathbf{T}_1}\left(\mathbf{t}_1(\mathbf{x});\mathcal{H}_1\right)}{p_{\mathbf{T}_1}\left(\mathbf{t}_1(\mathbf{x});\mathcal{H}_0\right)}\right)^{\eta_1}\left(\frac{p_{\mathbf{T}_2}\left(\mathbf{t}_2(\mathbf{x});\mathcal{H}_1\right)}{p_{\mathbf{T}_2}\left(\mathbf{t}_2(\mathbf{x});\mathcal{H}_0\right)}\right)^{\eta_2}\\&\times p_{\mathbf{X}}\left(\mathbf{x};H_0\right)d\mathbf{x}\end{aligned}\tag{3}$$

$$l_{\mathbf{T}_1}(\mathbf{x})=\ln\frac{p_{\mathbf{T}_1}\left(\mathbf{t}_1(\mathbf{x});\mathcal{H}_1\right)}{p_{\mathbf{T}_1}\left(\mathbf{t}_1(\mathbf{x});\mathcal{H}_0\right)},\ l_{\mathbf{T}_2}(\mathbf{x})=\ln\frac{p_{\mathbf{T}_2}\left(\mathbf{t}_2(\mathbf{x});\mathcal{H}_1\right)}{p_{\mathbf{T}_2}\left(\mathbf{t}_2(\mathbf{x});\mathcal{H}_0\right)}\tag{4}$$

(1) can be written as

$$\begin{aligned}p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta})=\exp&\left[\eta_1 l_{\mathbf{T}_1}(\mathbf{x})+\eta_2 l_{\mathbf{T}_2}(\mathbf{x})-K(\eta)\right.\\&\left.+\ln p_{\mathbf{X}}\left(\mathbf{x};H_0\right)\right]\end{aligned}\tag{5}$$

which is a two-parameter exponential family [9]. $K(\boldsymbol{\eta})$ is recognized as the cumulant generating function of $l_{\mathbf{T}_1}(\mathbf{x})$, $l_{\mathbf{T}_2}(\mathbf{x})$ when the PDF of $\mathbf{x}$ is $p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)$. Since (5) is of an exponential family, the EEF inherits some useful properties that we will discuss in the following (refer to [9], [10] and [11] for details).

1) If the PDF of $\mathbf{x}$ is $p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta})$, then the joint PDF of $\mathbf{T}_1$ and $\mathbf{T}_2$ is [11]

$$\begin{aligned}p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\boldsymbol{\eta})=\exp&\left[\eta_1 l_{\mathbf{T}_1}+\eta_2 l_{\mathbf{T}_2}-K(\boldsymbol{\eta})\right.\\&\left.+\ln p_{\mathbf{T}_1,\mathbf{T}_2}(\mathbf{t}_1,\mathbf{t}_2;\mathcal{H}_0)\right]\end{aligned}\tag{6}$$

where

$$l_{\mathbf{T}_1}=\ln\frac{p_{\mathbf{T}_1}\left(\mathbf{t}_1;\mathcal{H}_1\right)}{p_{\mathbf{T}_1}\left(\mathbf{t}_1;\mathcal{H}_0\right)},\ l_{\mathbf{T}_2}=\ln\frac{p_{\mathbf{T}_2}\left(\mathbf{t}_2;\mathcal{H}_1\right)}{p_{\mathbf{T}_2}\left(\mathbf{t}_2;\mathcal{H}_0\right)}\tag{7}$$

This can also be easily proved using surface integral techniques [12]. Notice that for (6), $\mathbf{T}_1$ and $\mathbf{T}_2$ are not independent unless they are independent under $\mathcal{H}_0$.

2) $K(\boldsymbol{\eta})$ is convex by Holder's inequality [9]. If we assume that $l_{\mathbf{T}_1}$ and $l_{\mathbf{T}_2}$ are linearly independent [13], then $\boldsymbol{\eta}$ is identifiable, and hence $K(\boldsymbol{\eta})$ is strictly convex [10].

3) Let $E_{\boldsymbol{\eta}}(l_{\mathbf{T}_i})$ be the expected value of $l_{\mathbf{T}_i}$ for $i=1,2$ and $\mathbf{C}_{\boldsymbol{\eta}}$ be the covariance matrix of $\left[l_{\mathbf{T}_1},\ l_{\mathbf{T}_2}\right]^T$ when $\mathbf{x}$ is distributed according to $p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta})$. We have

$$\frac{\partial K(\eta)}{\partial\eta_i}=E_\eta\left(l_{\mathbf{T}_i}\right)\tag{8}$$

$$\begin{bmatrix}\frac{\partial^2K(\boldsymbol{\eta})}{\partial\eta_1^2}&\frac{\partial^2K(\boldsymbol{\eta})}{\partial\eta_1\partial\eta_2}\\\frac{\partial^2K(\boldsymbol{\eta})}{\partial\eta_2\partial\eta_1}&\frac{\partial^2K(\boldsymbol{\eta})}{\partial\eta_2^2}\end{bmatrix}=\mathbf{C}_{\boldsymbol{\eta}}\tag{9}$$

Notice that (9) also shows that $K(\boldsymbol{\eta})$ is convex.

4) $\left[l_{\mathbf{T}_1},\ l_{\mathbf{T}_2}\right]^T$ is a minimal and complete sufficient statistic for $\boldsymbol{\eta}$. Hence $\left[l_{\mathbf{T}_1},\ l_{\mathbf{T}_2}\right]^T$ can be used to discriminate between $p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_1)$ and $p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)$.

5) $K(\boldsymbol{\eta})$ is finite on S. To see this, $K(\boldsymbol{\eta})>-\infty$ by definition. Obviously, $K(\boldsymbol{\eta})=0$ for $\boldsymbol{\eta}=[0,0]^T,[1,0]^T,[0,1]^T$. Since $K(\boldsymbol{\eta})$ is strictly convex, we have $K(\boldsymbol{\eta})\le 0<\infty$ for $\boldsymbol{\eta}\in S$. But when $\boldsymbol{\eta}$ is outside S, there is no guarantee that $K(\boldsymbol{\eta})$ is finite in general. This explains why we have the constraints in (2).

3. EEF FOR HYPOTHESIS TESTING

For binary hypothesis testing, we will decide $\mathcal{H}_1$ if

$$\max_{\boldsymbol{\eta}}\ln\frac{p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta})}{p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)}>\tau\tag{10}$$

where τ is a threshold. This test statistic actually does not depend on $\mathbf{x}$ but only on $\mathbf{t}_1$ and $\mathbf{t}_2$ since

$$g(\boldsymbol{\eta})=\ln\frac{p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta})}{p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)}=\eta_1 l_{\mathbf{T}_1}+\eta_2 l_{\mathbf{T}_2}-K(\boldsymbol{\eta})\tag{11}$$

The reason why we choose this test statistic, as we will show next, is that asymptotically $\max_{\boldsymbol{\eta}} p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta})$ is the closest to the unknown $p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_1)$ in KL divergence.

Assume that there are a large number of independent and identically distributed (IID) unobservable $\mathbf{x}_i$'s for $i = 1, 2, \ldots, N$, which results in IID $\mathbf{t}_{1i}$'s and IID $\mathbf{t}_{2i}$'s. We want to maximize

$$\begin{aligned} & \frac{1}{N}\sum_{i=1}^{N} \ln \frac{p_{\mathbf{X}}(\mathbf{x}_i;\boldsymbol{\eta})}{p_{\mathbf{X}}(\mathbf{x}_i;\mathcal{H}_0)} \\ &= \exp\left[\eta_1 \frac{1}{N}\sum_{i=1}^{N} l_{\mathbf{T}_{1i}} + \eta_2 \frac{1}{N}\sum_{i=1}^{N} l_{\mathbf{T}_{2i}} - K(\boldsymbol{\eta})\right] \end{aligned} \tag{12}$$

By the law of large number, under $\mathcal{H}_1$

$$\frac{1}{N}\sum_{i=1}^{N} l_{\mathbf{T}_{1i}} \to E_{\mathcal{H}_1}\left(l_{\mathbf{T}_1}\right) = D\left(p_{\mathbf{T}_1}\left(\mathbf{t}_1;\mathcal{H}_1\right) \| p_{\mathbf{T}_1}\left(\mathbf{t}_1;\mathcal{H}_0\right)\right)$$

$$\frac{1}{N}\sum_{i=1}^{N} l_{\mathbf{T}_{2i}} \to E_{\mathcal{H}_1}\left(l_{\mathbf{T}_2}\right) = D\left(p_{\mathbf{T}_2}\left(\mathbf{t}_2;\mathcal{H}_1\right) \| p_{\mathbf{T}_2}\left(\mathbf{t}_2;\mathcal{H}_0\right)\right)$$

as $N \to \infty$. So we are asymptotically maximizing

$$\begin{aligned} & \eta_1 D\left(p_{\mathbf{T}_1}\left(\mathbf{t}_1;\mathcal{H}_1\right) \| p_{\mathbf{T}_1}\left(\mathbf{t}_1;\mathcal{H}_0\right)\right) \\ & + \eta_2 D\left(p_{\mathbf{T}_2}\left(\mathbf{t}_2;\mathcal{H}_1\right) \| p_{\mathbf{T}_2}\left(\mathbf{t}_2;\mathcal{H}_0\right)\right) - K(\boldsymbol{\eta}) \end{aligned} \tag{13}$$

Since

$$\ln \frac{p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_1)}{p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta})} = -\eta_1 l_{\mathbf{T}_1} - \eta_2 l_{\mathbf{T}_2} + K(\boldsymbol{\eta}) + \ln \frac{p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_1)}{p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)}$$

the KL divergence between $p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_1)$ and $p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta})$ is

$$\begin{aligned} & D\left(p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_1) \| p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta})\right) \\ &= E_{\mathcal{H}_1} \exp\left[-\eta_1 l_{\mathbf{T}_1} - \eta_2 l_{\mathbf{T}_2} + K(\boldsymbol{\eta}) + \ln \frac{p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_1)}{p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_0)}\right] \\ &= -\eta_1 D\left(p_{\mathbf{T}_1}\left(\mathbf{t}_1;\mathcal{H}_1\right) \| p_{\mathbf{T}_1}\left(\mathbf{t}_1;\mathcal{H}_0\right)\right) \\ &\quad - \eta_2 D\left(p_{\mathbf{T}_2}\left(\mathbf{t}_2;\mathcal{H}_1\right) \| p_{\mathbf{T}_2}\left(\mathbf{t}_2;\mathcal{H}_0\right)\right) \\ &\quad + K(\boldsymbol{\eta}) + D\left(p_{\mathbf{X}}\left(\mathbf{x};\mathcal{H}_1\right) \| p_{\mathbf{X}}\left(\mathbf{x};\mathcal{H}_0\right)\right) \end{aligned} \tag{14}$$

This shows that $D\left(p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_1) \| p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta})\right)$ is minimized by maximizing (13). A similar result is shown in [6] by using a Pythagorean-like theorem. Also if $\mathbf{T}_1$ and/or $\mathbf{T}_2$ are sufficient statistics for deciding between $\mathcal{H}_0$ and $\mathcal{H}_1$, it can be shown that $p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta}) = p_{\mathbf{X}}(\mathbf{x};\mathcal{H}_1)$. Thus, the true PDF under $\mathcal{H}_1$ is recovered [14].

To implement (10), we require the maximum likelihood estimate (MLE) of $\boldsymbol{\eta}$. Let $\boldsymbol{\eta}^*$ be the MLE of $\boldsymbol{\eta}$ without constraints in (2). Since $g(\boldsymbol{\eta})$ is strictly concave, $\boldsymbol{\eta}^*$ is unique. Taking partial derivatives of $g(\boldsymbol{\eta})$ and setting to 0, we have

$$l_{\mathbf{T}_1} = \frac{\partial K(\boldsymbol{\eta})}{\partial \eta_1}\Big|_{\boldsymbol{\eta}^*}, \; l_{\mathbf{T}_2} = \frac{\partial K(\boldsymbol{\eta})}{\partial \eta_2}\Big|_{\boldsymbol{\eta}^*} \tag{15}$$

Let $\hat{\boldsymbol{\eta}}$ be the MLE of $\boldsymbol{\eta}$ with the constraints. If $\boldsymbol{\eta}^*$ is in the constraint set S, then $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}^*$. Otherwise, $\hat{\boldsymbol{\eta}}$ is unique and is on the boundary of S since $-g(\boldsymbol{\eta})$ is strictly convex and S is convex also [15], and hence we could simply search the boundary of S to find $\hat{\boldsymbol{\eta}}$.

4. EXAMPLE

Since only $\mathbf{T}_1$ and $\mathbf{T}_2$ are used in hypothesis testing, we only need to specify their distributions. Consider the case when $\mathbf{T}_1$ and $\mathbf{T}_2$ are scalars (will write them as T_1 and T_2) with distributions

$$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{bmatrix}\right) \quad \text{under } \mathcal{H}_0$$

$$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}\right) \quad \text{under } \mathcal{H}_1$$

where ρ_0 is known but ρ_1 *is unknown (we do not need the joint PDF of* T_1 *and* T_2 *under* $\mathcal{H}_1$*)*. We have

$$\begin{aligned} K(\boldsymbol{\eta}) &= \ln E_{\mathcal{H}_0}\left[\exp\left(\eta_1 l_{T_1} + \eta_2 l_{T_2}\right)\right] \\ &= \ln E_{\mathcal{H}_0}\left[\exp\left(\eta_1 \frac{2t_1A_1 - A_1^2}{2\sigma^2} + \eta_2 \frac{2t_2A_2 - A_2^2}{2\sigma^2}\right)\right] \\ &= -\eta_1 \frac{A_1^2}{2\sigma^2} - \eta_2 \frac{A_2^2}{2\sigma^2} \\ &\quad + \ln E_{\mathcal{H}_0}\left[\exp\left(\frac{\eta_1 t_1 A_1 + \eta_2 t_2 A_2}{\sigma^2}\right)\right] \end{aligned}$$

Let $\boldsymbol{\phi} = \left[\frac{\eta_1 A_1}{\sigma^2}, \; \frac{\eta_2 A_2}{\sigma^2}\right]^T$ and $\mathbf{t} = [t_1, \; t_2]^T$, then

$$\begin{aligned} E_{\mathcal{H}_0}\left[\exp\left(\frac{\eta_1 t_1 A_1 + \eta_2 t_2 A_2}{\sigma^2}\right)\right] &= E_{\mathcal{H}_0}\left[\exp\left(\boldsymbol{\phi}^T \mathbf{t}\right)\right] \\ &= \exp\left(\frac{1}{2}\boldsymbol{\phi}^T \mathbf{C}_0 \boldsymbol{\phi}\right) \end{aligned}$$

where $\mathbf{C}_0 = \sigma^2 \begin{bmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{bmatrix}$ and hence

$$K(\boldsymbol{\eta}) = -\eta_1 \frac{A_1^2}{2\sigma^2} - \eta_2 \frac{A_2^2}{2\sigma^2} + \frac{1}{2}\boldsymbol{\phi}^T \mathbf{C}_0 \boldsymbol{\phi}$$

So

$$\begin{aligned} g(\boldsymbol{\eta}) &= \eta_1 l_{T_1} + \eta_2 l_{T_2} - K(\boldsymbol{\eta}) \\ &= \eta_1 \frac{2t_1A_1 - A_1^2}{2\sigma^2} + \eta_2 \frac{2t_2A_2 - A_2^2}{2\sigma^2} - K(\boldsymbol{\eta}) \\ &= \frac{\eta_1 A_1 t_1}{\sigma^2} + \frac{\eta_2 A_2 t_2}{\sigma^2} - \frac{1}{2}\boldsymbol{\phi}^T \mathbf{C}_0 \boldsymbol{\phi} \\ &= \mathbf{t}^T \boldsymbol{\phi} - \frac{1}{2}\boldsymbol{\phi}^T \mathbf{C}_0 \boldsymbol{\phi} \end{aligned}$$

Differentiating and setting to 0, the global maximum is found at

$$\boldsymbol{\phi}^* = \mathbf{C}_0^{-1}\mathbf{t} = \begin{bmatrix} \frac{t_1 - \rho_0 t_2}{1-\rho_0^2} \\ \frac{t_2 - \rho_0 t_1}{1-\rho_0^2} \end{bmatrix}$$

or

$$\boldsymbol{\eta}^* = \begin{bmatrix} \frac{\sigma^2 (t_1 - \rho_0 t_2)}{A_1 (1-\rho_0^2)} \\ \frac{\sigma^2 (t_2 - \rho_0 t_1)}{A_2 (1-\rho_0^2)} \end{bmatrix}$$

If $\boldsymbol{\eta}^* \in S$, then we decide $\mathcal{H}_1$ if $g(\boldsymbol{\eta}^*) = \mathbf{t}^T\mathbf{C}_0^{-1}\mathbf{t} > \tau$, otherwise we search $\hat{\boldsymbol{\eta}}$ on the boundary and decide $\mathcal{H}_1$ if $g(\hat{\boldsymbol{\eta}}) > \tau$.

When we observe N IID t_{1i}'s and IID t_{2i}'s, then it easily extends that by (12), $[t_1,\ t_2]^T$ is replaced by the sample mean $\left[\frac{1}{N}\sum_{i=1}^N t_{1i},\ \frac{1}{N}\sum_{i=1}^N t_{2i}\right]^T$, and everything else remains the same.

5. SIMULATION RESULTS

For the above example, we set $N = 20$, $A_1 = 0.3$, $A_2 = 0.35$, $\sigma^2 = 1$, $\rho_0 = 0.6$ and $\rho_1 = 0.7$. We compare the EEF approach with the clairvoyant detector (ρ_1 is known, its performance is an upper bound), the detector assuming independence of t_1 and t_2, and the copula based method. The copula method estimates the linear correlation coefficient ρ_1 using a non-parametric rank correlation measure, Kendall's τ. We use the Gaussian copula as in [3]. The simulation is repeated for 5000 trials. The receiver operating characteristic curves (ROC) are plotted. As seen in Fig. 1, the EEF is only poorer than the clairvoyant detector, and performs better than the other two methods.

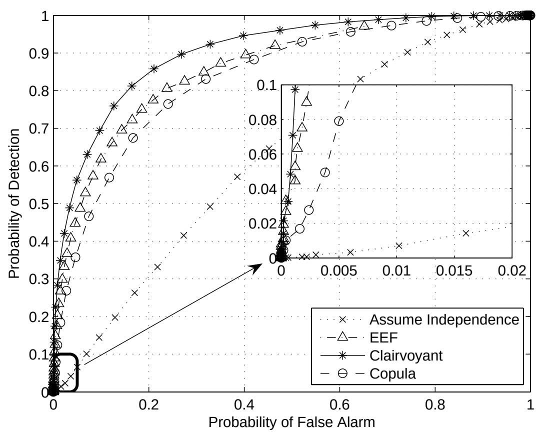

Fig. 1. ROC curves for different detectors.

6. CONCLUSION

The EEF based approach is proposed for the problem of multimodal signal processing when the outputs are not independent. It exponentially embeds two or more PDFs and approximates an unknown PDF. Such embedding is highly related to the KL divergence and many of its properties have been discussed. Examples are given to help understand the application of this method. Compared to some existing approaches, better performance is observed for the proposed method. The connections among $\hat{\boldsymbol{\eta}}$, $K(\boldsymbol{\eta})$ and the KL divergence and more of its theoretical properties will be investigated in the future.

7. REFERENCES

[1] S.C.A. Thomopoulos, R. Viswanathan, and D.K. Bougoulias, "Optimal distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, pp. 761–765, Sep. 1989.

[2] Z. Chair and P.K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 98–101, Jan. 1986.

[3] A. Sundaresan, P.K. Varshney, and N.S.V. Rao, "Distributed detection of a nuclear radioactive source using fusion of correlated decisions," in *Information Fusion, 2007 10th International Conference on*, 2007, pp. 1–7.

[4] S.G. Iyengar, P.K. Varshney, and T. Damarla, "A parametric copula based framework for multimodal signal processing," in *ICASSP*, 2009, pp. 1893–1896.

[5] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley and Sons, second edition, 2006.

[6] S. Kay, "Exponentially embedded families - new approaches to model order estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, pp. 333–345, Jan. 2005.

[7] S. Kullback, *Information Theory and Statistics*, Courier Dover Publications, second edition, 1997.

[8] S. Kay, A. Nuttall, and P.M. Baggenstoss, "Multidimensional probability density function approximations for detection, classification, and model order selection," *IEEE Trans. Signal Process.*, vol. 49, pp. 2240–2252, Oct. 2001.

[9] L.D. Brown, *Fundamentals of Statistical Exponential Families*, Institute of Mathematical Statistics, 1986.

[10] P.J. Bickel and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, vol. 1, Pearson Prentice Hall, 2006.

[11] E.L. Lehmann and J.P. Romano, *Testing Statistical Hypotheses*, Springer, third edition, 2005.

[12] J.J. Higgins, "Some surface integral techniques in statistics," *The American Statistician*, vol. 29, pp. 43–46, Feb. 1975.

[13] J. Pfanzagl and W. Wefelmeyer, *Contributions to a General Asymptotic Statistical Theory*, vol. 13 of *Lecture Notes in Statistics*, Springer-Verlag, 1982.

[14] S. Kay, "Asymptotically optimal approximation of multidimensional pdf's by lower dimensional pdf's," *IEEE Trans. Signal Process.*, vol. 55, pp. 725–729, Feb. 2007.

[15] D.G. Luenberger, *Linear and Nonlinear Programming*, Springer, second edition, 2003.