MEASURE TRANSFORMED QUASI LIKELIHOOD RATIO TEST FOR BAYESIAN BINARY HYPOTHESIS TESTING

Nir Halay, Koby Todros

Ben-Gurion University of the Negev

Alfred O. Hero

University of Michigan

ABSTRACT

In this paper, a generalization of the Gaussian quasi likelihood ratio test (GQLRT) for Bayesian binary hypothesis testing is developed. The proposed generalization, called measure-transformed GQLRT (MT-GQLRT), selects a Gaussian probability model that best empirically fits a transformed conditional probability measure of the data. By judicious choice of the transform we show that, unlike the GQLRT, the proposed test is resilient to outliers and involves higher-order statistical moments leading to significant mitigation of the model mismatch effect on the decision performance. Under some mild regularity conditions we show that the test statistic of the proposed MT-GQLRT is asymptotically normal. A data driven procedure for optimal selection of the measure transformation parameters is developed that minimizes an empirical estimate of the asymptotic Bayes risk. The MT-GQLRT is applied to signal classification in a simulation example that establishes significantly improved probability of error performance relative to the standard GQLRT.

***Index Terms*—** Bayesian hypothesis testing, Higher-order statistics, probability measure transform, robust statistics, signal classification.

1. INTRODUCTION

Bayesian binary hypothesis testing deals with deciding between two hypotheses based on a sequence of multivariate samples from an underlying probability distribution that is equal to one of two conditional probability measures [1]. When the prior probabilities of the considered hypotheses are known and the conditional probability distributions under each hypothesis are correctly specified the likelihood ratio test (LRT), which minimizes the Bayes risk [1], can be implemented. In many practical scenarios the conditional probability distributions are unknown, and therefore, one must resort to suboptimal tests.

A popular test of this kind is the Bayesian Gaussian quasi LRT (GQLRT) [2]-[5] which assumes that the probability distributions of the samples conditioned on each hypothesis are Gaussian. The Bayesian GQLRT operates by selecting the Gaussian conditional probability model that best fits the data. When the observations are i.i.d. this selection is carried out by comparing the empirical Kullback-Leibler divergences [6] between the underlying conditional probability distribution and the assumed normal probability measures. The Bayesian GQLRT has gained popularity due to its implementation simplicity, ease of performance analysis, and its geometrical interpretations. However, in some circumstances, such as for certain types of non-Gaussian data, deviation from normality can degrade decision performance. This can occur when the first and second-order statistical moments are weakly identifiable over the considered hypotheses, or in the case of heavy-tailed data when the non-robust sample mean and covariance are sensitive to outliers.

In this paper, a generalization of the Bayesian GQLRT is proposed that operates by selecting a Gaussian probability model that has the best empirical fit to a transformed conditional probability distribution of the data. The proposed test is a Bayesian version of the test proposed in [7] for non-Bayesian binary hypothesis testing. Under the proposed generalization outliers-resilient tests can be obtained that involve higher-order statistical moments, and yet have the computational and implementation advantages of the Bayesian GQLRT.

The proposed transform is structured by a non-negative function, called the MT-function, and maps the conditional probability distribution into a set of new conditional probability measures on the observation space. By modifying the MT-function, classes of measure transformations can be obtained that have different useful properties. Under the proposed transform we define the measure-transformed (MT) mean vector and covariance matrix, derive their strongly consistent estimates, and study their relation to higher-order statistical moments and resilience to outliers.

Similarly to the Bayesian GQLRT, the proposed MT-GQLRT compares the empirical Kullback-Leibler divergences between the transformed conditional probability distribution of the data and two normal probability measures that are characterized by the MT-mean vector and MT-covariance matrix conditioned on each hypothesis. Under some mild regularity conditions we show that the proposed test statistic is asymptotically normal. Furthermore, given two training sequences from the conditional probability distribution of each hypothesis, a data-driven procedure for optimal selection of the MT-function within some parametric class of functions is developed that minimizes an empirical estimate of the Bayes risk.

We illustrate the MT-GQLRT for the problem of Bayesian signal classification in the presence of heavy-tailed spherically contoured noise [8] that produces outliers. By specifying the MT-function within the family of zero-centered Gaussian functions parameterized by a scale parameter, we show that the MT-GQLRT outperforms the non-robust Bayesian GQLRT and attains classification performance that are significantly closer to those obtained by the omniscient Bayesian LRT that, unlike the MT-GQLRT, requires complete knowledge of the conditional likelihood function given each hypothesis.

The paper is organized as follows. In Section 2, the principles of our proposed probability measure transform are reviewed. In Section 3, we use this transformation to construct the MT-GQLRT. The proposed test is applied to a signal classification problem in Section 4. In Section 5, the main points of this contribution are summarized. Proofs for the theorems, propositions and corollaries stated throughout the paper will be provided in the full length journal version.

This work was partially supported by United States-Israel Binational Science Foundation grant 2014334 and United States ARO MURI grant W911NF-11-1-0391.

2. PROBABILITY MEASURE TRANSFORM

In this section, we develop a transform on the conditional probability measure of the data. Under the proposed transform, we define the conditional measure-transformed mean vector and covariance matrix, derive their strongly consistent estimates, and establish their relation to higher-order statistical moments and resilience to outliers. These quantities will be used in the following section to construct the proposed measure-transformed GQLRT.

2.1. Probability measure transformation

We define the measure space $\left(\mathcal{X}, \mathcal{S}_{\mathcal{X}}, P_{\mathbf{X}|\boldsymbol{\theta}}\right)$, where $\mathcal{X} \subseteq \mathbb{C}^p$ is the observation space of a continuous random vector $\mathbf{X}$, $\mathcal{S}_{\mathcal{X}}$ is a σ-algebra over $\mathcal{X}$ and $P_{\mathbf{X}|\boldsymbol{\theta}}$ is an unknown probability measure on $\mathcal{S}_{\mathcal{X}}$ conditioned on a random vector $\boldsymbol{\theta}$ that takes values in the pair set $\boldsymbol{\Theta} \triangleq \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$ with known a-priori probabilities, $P_{\boldsymbol{\theta}_0}$ and $P_{\boldsymbol{\theta}_1}$, respectively.

Definition 1. *Given a non-negative function $u : \mathbb{C}^p \to \mathbb{R}_+$ satisfying*

$$0 < \mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X}|\boldsymbol{\theta}}\right] < \infty \qquad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}, \tag{1}$$

where $\mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X}|\boldsymbol{\theta}}\right] \triangleq \int_{\mathcal{X}} u\left(\mathbf{x}\right) dP_{\mathbf{X}|\boldsymbol{\theta}}\left(\mathbf{x}\right)$ and $\mathbf{x} \in \mathcal{X}$, a transform on $P_{\mathbf{X}|\boldsymbol{\theta}}$ is defined via the relation:

$$Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}\left(A\right) \triangleq \mathrm{T}_u\left[P_{\mathbf{X}|\boldsymbol{\theta}}\right]\left(A\right) = \int_A \varphi_u\left(\mathbf{x}; \boldsymbol{\theta}\right) dP_{\mathbf{X}|\boldsymbol{\theta}}\left(\mathbf{x}\right), \tag{2}$$

where $A \in \mathcal{S}_{\mathcal{X}}$ and $\varphi_u\left(\mathbf{x}; \boldsymbol{\theta}\right) \triangleq u\left(\mathbf{x}\right) / \mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X}|\boldsymbol{\theta}}\right]$. The function $u\left(\cdot\right)$ is called the MT-function.

Proposition 1 (Properties of the transform)**.** *Let $Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ be defined by relation (2). Then 1) $Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ is a probability measure on $\mathcal{S}_{\mathcal{X}}$. 2) $Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ is absolutely continuous w.r.t. $P_{\mathbf{X}|\boldsymbol{\theta}}$, with Radon-Nikodym derivative [9]:*

$$dQ^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}\left(\mathbf{x}\right) / dP_{\mathbf{X}|\boldsymbol{\theta}}\left(\mathbf{x}\right) = \varphi_u\left(\mathbf{x}; \boldsymbol{\theta}\right). \tag{3}$$

By modifying $u\left(\cdot\right)$, such that the condition (1) is satisfied, virtually any probability measure on $\mathcal{S}_{\mathcal{X}}$ can be obtained.

2.2. The conditional MT-mean and MT-covariance

According to (3) the mean vector and covariance matrix of $\mathbf{X}$ under the transformed conditional distribution $Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ are given by:

$$\boldsymbol{\mu}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}} \triangleq \mathrm{E}\left[\mathbf{X}\varphi_u\left(\mathbf{X}; \boldsymbol{\theta}\right); P_{\mathbf{X}|\boldsymbol{\theta}}\right] \tag{4}$$

and

$$\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}} \triangleq \mathrm{E}\left[\mathbf{X}\mathbf{X}^H\varphi_u\left(\mathbf{X}; \boldsymbol{\theta}\right); P_{\mathbf{X}|\boldsymbol{\theta}}\right] - \boldsymbol{\mu}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}\boldsymbol{\mu}^{(u)H}_{\mathbf{X}|\boldsymbol{\theta}}, \tag{5}$$

respectively. Equations (4) and (5) imply that $\boldsymbol{\mu}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ are weighted mean and covariance of $\mathbf{X}$ under $P_{\mathbf{X}|\boldsymbol{\theta}}$, with the weighting function $\varphi_u\left(\cdot;\cdot\right)$ defined below (2). By choosing $u\left(\cdot\right)$ to be any non-zero constant valued function we have $Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}} = P_{\mathbf{X}|\boldsymbol{\theta}}$, for which the standard conditional mean vector $\boldsymbol{\mu}_{\mathbf{X}|\boldsymbol{\theta}}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}|\boldsymbol{\theta}}$ are obtained. Alternatively, when $u\left(\cdot\right)$ is non-constant analytic function, which has a convergent Taylor series expansion, the resulting conditional MT-mean and MT-covariance involve higher-order statistical moments of $P_{\mathbf{X}|\boldsymbol{\theta}}$.

2.3. Estimates of the conditional MT-mean and MT-covariance

Given a sequence of N i.i.d. samples from $P_{\mathbf{X}|\boldsymbol{\theta}}$ the empirical estimates of $\boldsymbol{\mu}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ are defined as:

$$\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} \triangleq \sum_{n=1}^{N} \mathbf{X}_n \hat{\varphi}_u\left(\mathbf{X}_n\right) \tag{6}$$

and

$$\hat{\boldsymbol{\Sigma}}^{(u)}_{\mathbf{X}} \triangleq \sum_{n=1}^{N} \mathbf{X}_n \mathbf{X}_n^H \hat{\varphi}_u\left(\mathbf{X}_n\right) - \hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{x}}\hat{\boldsymbol{\mu}}^{(u)H}_{\mathbf{x}}, \tag{7}$$

respectively, where $\hat{\varphi}_u\left(\mathbf{X}_n\right) \triangleq u\left(\mathbf{X}_n\right) / \sum_{n=1}^{N} u\left(\mathbf{X}_n\right)$. According to Proposition 2 in [10], if $\mathrm{E}\left[\|\mathbf{X}\|_2^2\, u\left(\mathbf{X}\right); P_{\mathbf{X}|\boldsymbol{\theta}}\right] < \infty$ then $\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} \xrightarrow[N\to\infty]{\text{w.p. 1}} \boldsymbol{\mu}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Sigma}}^{(u)}_{\mathbf{X}} \xrightarrow[N\to\infty]{\text{w.p. 1}} \boldsymbol{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$, where "$\xrightarrow{\text{w.p. 1}}$" denotes convergence with probability (w.p.) 1 [11].

Robustness of the empirical MT-covariance (7) to outliers was studied in [10] using its influence function [12] which describes the effect on the estimator of an infinitesimal contamination at some point $\mathbf{y} \in \mathbb{C}^p$. An estimator is said to be B-robust if its influence function is bounded [12]. Similarly to the proof of Proposition 3 in [10] it can be shown that if the MT-function $u(\mathbf{y})$ and the product $u(\mathbf{y})\|\mathbf{y}\|_2^2$ are bounded over $\mathbb{C}^p$ then the influence functions of both (6) and (7) are bounded.

3. MEASURE-TRANSFORMED GAUSSIAN QUASI LIKELIHOOD RATIO TEST

In this section we use the measure transformation (2) to construct a Bayesian test between the null and alternative hypotheses $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$ based on a sequence of samples $\mathbf{X}_n$, $n = 1, \ldots, N$ from the conditional distribution $P_{\mathbf{X}|\boldsymbol{\theta}}$. Regularity conditions for asymptotic normality of the corresponding test statistic are derived. Under these conditions, a consistent estimate of the asymptotic Bayes risk is derived. Optimal selection of the MT-function $u\left(\cdot\right)$ out of some parametric class of functions is also discussed.

3.1. The MT-GQLRT

Let $\Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ denote a complex circular Gaussian probability distribution [13] that is characterized by the conditional MT-mean $\boldsymbol{\mu}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ and MT-covariance $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$. The KLD between $Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ and $\Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_k}$, $k \in \{0, 1\}$ is defined as [6]:

$$D_{\mathrm{KL}}\left[Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}} || \Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_k}\right] \triangleq \mathrm{E}\left[\log \frac{q^{(u)}\left(\mathbf{X}|\boldsymbol{\theta}\right)}{\phi^{(u)}\left(\mathbf{X}|\boldsymbol{\theta}_k\right)}; Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}\right], \tag{8}$$

where $q^{(u)}\left(\mathbf{x}|\boldsymbol{\theta}\right)$ and $\phi^{(u)}\left(\mathbf{x}|\boldsymbol{\theta}_k\right)$ are the density functions of $Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}$ and $\Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_k}$, respectively. When $\Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_0} \neq \Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_1}$, the difference $D[Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}} || \Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_0}] - D[Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}} || \Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_1}]$ will be negative when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and positive when $\boldsymbol{\theta} = \boldsymbol{\theta}_1$. This motivates an empirical estimate of this difference as a test statistic for testing H_0 versus H_1. According to (3), $D_{\mathrm{KL}}\left[Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}} || \Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_k}\right]$ can be estimated using only samples from $P_{\mathbf{X}|\boldsymbol{\theta}}$. Hence, similarly to (6) and (7), an empirical estimate of (8) given a sequence of samples $\mathbf{X}_n$, $n = 1, \ldots, N$ from $P_{\mathbf{X}|\boldsymbol{\theta}}$ is defined as:

$$\hat{D}_{\mathrm{KL}}\left[Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}} || \Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_k}\right] \triangleq \sum_{n=1}^{N} \hat{\varphi}_u\left(\mathbf{X}_n\right) \log \frac{q^{(u)}\left(\mathbf{X}_n|\boldsymbol{\theta}\right)}{\phi^{(u)}\left(\mathbf{X}_n|\boldsymbol{\theta}_k\right)}, \tag{9}$$

where $\hat{\varphi}_u\left(\cdot\right)$ is defined below (7). Hence, we propose the following test statistic, which is independent of the unknown conditional density function $q^{(u)}(\mathbf{x}|\boldsymbol{\theta})$:

$$\begin{aligned} T_u &\triangleq \hat{D}_{\mathrm{KL}}\left[Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}||\Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_0}\right] - \hat{D}_{\mathrm{KL}}\left[Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}||\Phi^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_1}\right] \quad (10)\\ &= \sum_{n=1}^{N}\hat{\varphi}_u(\mathbf{X}_n)\,\psi_u(\mathbf{X}_n;\boldsymbol{\theta}_0,\boldsymbol{\theta}_1)\\ &= \left(D_{\mathrm{LD}}\left[\hat{\mathbf{\Sigma}}^{(u)}_{\mathbf{X}}||\mathbf{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_0}\right] + \left\|\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} - \boldsymbol{\mu}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_0}\right\|^2_{\left(\mathbf{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_0}\right)^{-1}}\right)\\ &- \left(D_{\mathrm{LD}}\left[\hat{\mathbf{\Sigma}}^{(u)}_{\mathbf{X}}||\mathbf{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_1}\right] + \left\|\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} - \boldsymbol{\mu}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_1}\right\|^2_{\left(\mathbf{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_1}\right)^{-1}}\right), \end{aligned}$$

where

$$\psi_u(\mathbf{X};\boldsymbol{\theta}_0,\boldsymbol{\theta}_1) \triangleq \log\frac{\phi^{(u)}(\mathbf{X}|\boldsymbol{\theta}_1)}{\phi^{(u)}(\mathbf{X}|\boldsymbol{\theta}_0)},$$

$D_{\mathrm{LD}}[\mathbf{A}||\mathbf{B}] \triangleq \mathrm{tr}\left[\mathbf{A}\mathbf{B}^{-1}\right] - \log\det\left[\mathbf{A}\mathbf{B}^{-1}\right] - p$ is the log-determinant divergence [14] between positive definite matrices $\mathbf{A},\mathbf{B}\in\mathbb{C}^{p\times p}$ and $\|\mathbf{a}\|_{\mathbf{C}} \triangleq \sqrt{\mathbf{a}^H\mathbf{C}\mathbf{a}}$ denotes the weighted Euclidian norm of a vector $\mathbf{a}\in\mathbb{C}^p$ with positive-definite weighting matrix $\mathbf{C}\in\mathbb{C}^{p\times p}$. The decision rule based on the test statistic (10) is given by

$$T_u \underset{H_0}{\overset{H_1}{\gtrless}} t, \tag{11}$$

where $t\in\mathbb{R}$ denotes a threshold value. Notice that for any non-zero constant MT-function, $u(\cdot)$, $Q^{(u)}_{\mathbf{X}|\boldsymbol{\theta}} = P_{\mathbf{X}|\boldsymbol{\theta}}$ and the standard Bayesian GQLRT is obtained, which only involves first and second-order moments.

3.2. Asymptotic performance analysis

Here, we study the asymptotic performance of the proposed test (11). We assume a sequence of i.i.d. samples $\mathbf{X}_n$, $n=1,\ldots,N$ from $P_{\mathbf{X}|\boldsymbol{\theta}}$.

Theorem 1 (Asymptotic normality)**.** *Assume that the following conditions are satisfied: 1)* $\boldsymbol{\mu}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_0} \neq \boldsymbol{\mu}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_1}$ *or* $\mathbf{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_0} \neq \mathbf{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_1}$*. 2)* $\mathbf{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_0}$ *and* $\mathbf{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}_1}$ *are non-singular. 3)* $\mathrm{E}\left[u^2(\mathbf{X});P_{\mathbf{X}|\boldsymbol{\theta}}\right]$ *and* $\mathrm{E}\left[\|\mathbf{X}\|_2^4\,u^2(\mathbf{X});P_{\mathbf{X}|\boldsymbol{\theta}}\right]$ *are finite for* $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ *and* $\boldsymbol{\theta}=\boldsymbol{\theta}_1$*. Then,*

$$\frac{T_u - \eta^{(u)}_{\boldsymbol{\theta}}}{\sqrt{\lambda^{(u)}_{\boldsymbol{\theta}}}} \xrightarrow[N\to\infty]{D} \mathcal{N}(0,1) \quad \forall\boldsymbol{\theta}\in\mathbf{\Theta},$$

where "$\xrightarrow{D}$*" denotes convergence in distribution [11], the mean* $\eta^{(u)}_{\boldsymbol{\theta}} \triangleq \mathrm{E}\left[\varphi_u(\mathbf{X};\boldsymbol{\theta})\,\psi_u(\mathbf{X};\boldsymbol{\theta}_0,\boldsymbol{\theta}_1);P_{\mathbf{X}|\boldsymbol{\theta}}\right]$ *and the variance* $\lambda^{(u)}_{\boldsymbol{\theta}} \triangleq N^{-1}\mathrm{E}\left[\varphi^2_u(\mathbf{X};\boldsymbol{\theta})\left(\psi_u(\mathbf{X};\boldsymbol{\theta}_0,\boldsymbol{\theta}_1) - \eta^{(u)}_{\boldsymbol{\theta}}\right)^2;P_{\mathbf{X}|\boldsymbol{\theta}}\right]$.

Corollary 1 (Asymptotic Bayes risk)**.** *Assume that the conditions stated in Theorem 1 are satisfied. For a loss function,*

$$L(\boldsymbol{\theta},\hat{\boldsymbol{\theta}}) = \begin{cases} L_{10}, & \text{if } \hat{\boldsymbol{\theta}}=\boldsymbol{\theta}_1 \text{ and } \boldsymbol{\theta}=\boldsymbol{\theta}_0\\ L_{01}, & \text{if } \hat{\boldsymbol{\theta}}=\boldsymbol{\theta}_0 \text{ and } \boldsymbol{\theta}=\boldsymbol{\theta}_1\\ 0, & \text{otherwise}\end{cases}, \tag{12}$$

where $\hat{\boldsymbol{\theta}}$ *denotes the outcome of the test (11), the asymptotic Bayes risk can be written as:*

$$R^{(u)}(t) \triangleq L_{10}P_{\boldsymbol{\theta}_0}Q\left(\frac{t-\eta^{(u)}_{\boldsymbol{\theta}_0}}{\sqrt{\lambda^{(u)}_{\boldsymbol{\theta}_0}}}\right) + L_{01}P_{\boldsymbol{\theta}_1}Q\left(\frac{\eta^{(u)}_{\boldsymbol{\theta}_1}-t}{\sqrt{\lambda^{(u)}_{\boldsymbol{\theta}_1}}}\right), \tag{13}$$

where $Q(\cdot)$ *denotes the tail probability of the standard normal distribution.*

In the following Proposition, a strongly consistent estimate of the asymptotic Bayes risk (13) is constructed based two i.i.d. sequences from $P_{\mathbf{X}|\boldsymbol{\theta}_0}$ and $P_{\mathbf{X}|\boldsymbol{\theta}_1}$. This quantity will be used in the sequel for optimal selection of the MT-function.

Proposition 2 (Empirical asymptotic Bayes risk)**.** *Let* $\mathbf{X}^{(k)}_n$*,* $n=1,\ldots,N_k$*,* $k=0,1$ *denote sequences of i.i.d. samples from* $P_{\mathbf{X}|\boldsymbol{\theta}_0}$ *and* $P_{\mathbf{X}|\boldsymbol{\theta}_1}$*, respectively. Define the empirical asymptotic Bayes risk:*

$$\hat{R}^{(u)}(t) \triangleq L_{10}P_{\boldsymbol{\theta}_0}Q\left(\frac{t-\hat{\eta}^{(u)}_{\boldsymbol{\theta}_0}}{\sqrt{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_0}}}\right) + L_{01}P_{\boldsymbol{\theta}_1}Q\left(\frac{\hat{\eta}^{(u)}_{\boldsymbol{\theta}_1}-t}{\sqrt{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_1}}}\right), \tag{14}$$

where $\hat{\eta}^{(u)}_{\boldsymbol{\theta}_k} \triangleq \sum_{n=1}^{N_k}\hat{\varphi}_u\left(\mathbf{X}^{(k)}_n\right)\psi_u\left(\mathbf{X}^{(k)}_n;\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right)$ *and* $\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_k} \triangleq \frac{N_k}{N}\sum_{n=1}^{N_k}\hat{\varphi}^2_u\left(\mathbf{X}^{(k)}_n\right)\left(\psi_u\left(\mathbf{X}^{(k)}_n;\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right) - \hat{\eta}^{(u)}_{\boldsymbol{\theta}_k}\right)^2$*. Assume that* $\mathrm{E}\left[u^2(\mathbf{X});P_{\mathbf{X}|\boldsymbol{\theta}}\right]$ *and* $\mathrm{E}\left[\|\mathbf{X}\|^4\,u^2(\mathbf{X});P_{\mathbf{X}|\boldsymbol{\theta}}\right]$ *are finite for any* $\boldsymbol{\theta}\in\mathbf{\Theta}$*. Then,* $\hat{R}^{(u)} \xrightarrow[N_0,N_1\to\infty]{w.p.1} R^{(u)}$.

It can be shown that the optimal threshold that minimizes (14) is:

$$t^{(u)}_{opt} \triangleq \frac{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_0}\hat{\eta}^{(u)}_{\boldsymbol{\theta}_1} - \hat{\lambda}^{(u)}_{\boldsymbol{\theta}_1}\hat{\eta}^{(u)}_{\boldsymbol{\theta}_0} - \sqrt{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_0}\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_1}\hat{s}^{(u)}}}{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_0} - \hat{\lambda}^{(u)}_{\boldsymbol{\theta}_1}}, \tag{15}$$

where $\hat{\lambda}^u_0 \neq \hat{\lambda}^u_1$ w.p. 1 since $\mathbf{X}$ is continuous random vector, and $\hat{s}^{(u)} \triangleq \left(\hat{\eta}^{(u)}_{\boldsymbol{\theta}_0} - \hat{\eta}^{(u)}_{\boldsymbol{\theta}_1}\right)^2 - 2\left(\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_0} - \hat{\lambda}^{(u)}_{\boldsymbol{\theta}_1}\right)\log\frac{L_{10}P_{\boldsymbol{\theta}_0}\sqrt{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_1}}}{L_{01}P_{\boldsymbol{\theta}_1}\sqrt{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_0}}}$ is non-negative when (14) has a global minimum with respect to t.

3.3. Optimal selection of the MT-function

We propose to specify the MT-function within some parametric family $\{u(\mathbf{X};\boldsymbol{\omega}),\boldsymbol{\omega}\in\mathbf{\Omega}\subseteq\mathbb{C}^r\}$ that satisfies the conditions stated in Definition 1 and Theorem 1. An optimal choice of the MT-function parameter $\boldsymbol{\omega}$ minimizes the empirical asymptotic Bayes risk (14) evaluated at the optimal threshold (15).

4. EXAMPLE

We consider the following Bayesian signal classification problem:

$$\begin{aligned} H_0 &: \mathbf{X}_n = S_n\boldsymbol{\theta}_0 + \mathbf{W}_n, \quad n=1,\ldots,N, \\ H_1 &: \mathbf{X}_n = S_n\boldsymbol{\theta}_1 + \mathbf{W}_n, \quad n=1,\ldots,N, \end{aligned} \tag{16}$$

where $\mathbf{X}_n\in\mathbb{C}^p$ is an observation vector, $S_n\in\mathbb{C}$ is a first-order stationary random signal that is symmetrically distributed about the origin, and $\boldsymbol{\theta}_0$,$\boldsymbol{\theta}_1$ are realizations of a binary unit norm random vector $\boldsymbol{\theta}$ with known a-priori probabilities $P_{\boldsymbol{\theta}_0}$ and $P_{\boldsymbol{\theta}_1}$, respectively. The vector $\mathbf{W}_n\in\mathbb{C}^p$ is a first-order stationary additive noise that is statistically independent of S_n. We assume that the noise component has a density that is spherically contoured with stochastic representation [8]:

$$\mathbf{W}_n = \nu_n\mathbf{Z}_n, \tag{17}$$

where $\nu_n\in\mathbb{R}_{++}$ is a first-order stationary process and $\mathbf{Z}_n\in\mathbb{C}^p$ is a proper-complex wide-sense stationary Gaussian process with zero-mean and scaled unit covariance $\sigma^2_{\mathbf{Z}}\mathbf{I}$. The processes ν_n and $\mathbf{Z}_n$ are assumed to be statistically independent.

In order to gain robustness against outliers, as well as sensitivity to higher-order moments, we specify the MT-function in the zero-centred Gaussian family of functions parametrized by a width parameter ω, i.e.,

$$u\left(\mathbf{x};\omega\right)=\exp\left(-\|\mathbf{x}\|^{2}/\omega^{2}\right),\quad \omega\in\mathbb{R}_{++}. \tag{18}$$

Notice that the MT-function (18) satisfies the B-robustness conditions stated at the ending paragraph of Subsection 2.3. Using (4), (5) and (16)-(18) it can be shown that the conditional MT-mean and MT-covariance are:

$$\boldsymbol{\mu}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}\left(\omega\right)=\mathbf{0}\quad\text{and}\quad\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}|\boldsymbol{\theta}}\left(\omega\right)=r_{S}\left(\omega\right)\boldsymbol{\theta}\boldsymbol{\theta}^{H}+r_{W}\left(\omega\right)\mathbf{I}, \tag{19}$$

respectively, where $r_S\left(\omega\right)$ and $r_W\left(\omega\right)$ are some strictly positive functions of ω. Hence, by substituting (18) and (19) into (10) followed by normalization by the observation-independent factor $c\left(\omega\right)\triangleq\frac{r_S(\omega)}{r_W(\omega)(r_S(\omega)+r_W(\omega))}$, the MT-GQLRT (11) simplifies to

$$T'_u\triangleq T_u/c(\omega)=\boldsymbol{\theta}_1^H\hat{\mathbf{C}}^{(u)}_{\mathbf{X}}\left(\omega\right)\boldsymbol{\theta}_1-\boldsymbol{\theta}_0^H\hat{\mathbf{C}}^{(u)}_{\mathbf{X}}\left(\omega\right)\boldsymbol{\theta}_0\underset{H_0}{\overset{H_1}{\gtrless}}t',$$

where $\hat{\mathbf{C}}^{(u)}_{\mathbf{X}}\left(\omega\right)\triangleq\hat{\boldsymbol{\Sigma}}^{(u)}_{\mathbf{X}}\left(\omega\right)+\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}}\left(\omega\right)\hat{\boldsymbol{\mu}}^{(u)H}_{\mathbf{X}}\left(\omega\right)$ and $t'\triangleq t/c(\omega)$.

Under the considered settings, it can be shown that the conditions stated in Proposition 1 are satisfied. We choose $L_{10}=L_{01}=1$, under which the asymptotic Bayes risk (13) reduces to the probability of error [1]. In this case it can be shown that the asymptotic minimum probability of error w.r.t. the threshold parameter takes the form:

$$P_e(\omega)=\sum_{k=0}^{1}P_{\boldsymbol{\theta}_k}Q\left(\frac{1}{H(\omega)}+(-1)^k\frac{1}{2}H(\omega)\log\frac{P_{\boldsymbol{\theta}_0}}{P_{\boldsymbol{\theta}_1}}\right), \tag{20}$$

where $H(\omega)\triangleq\sqrt{\frac{2G_1(\omega)+\left(1-|\boldsymbol{\theta}_0^H\boldsymbol{\theta}_1|^2\right)G_0(\omega)}{N\left(1-|\boldsymbol{\theta}_0^H\boldsymbol{\theta}_1|^2\right)}}$, $G_k(\omega)\triangleq\frac{\mathrm{E}\left[g_k\left(S,\sqrt{2}\bar{\nu},\omega\right)h\left(\sqrt{2}S,\sqrt{2}\bar{\nu},\omega,2k\right);P_{S,\nu}\right]}{\mathrm{E}^2\left[|S|^2h(S,\bar{\nu},\omega,2k);P_{S,\nu}\right]}$, $g_1(S,\nu,\omega)\triangleq\frac{\omega^2\nu^2|S|^2}{\omega^2+\nu^2}+\nu^4$, $g_0(S,\nu,\omega)\triangleq\left(\left(\frac{\omega^2|S|}{\omega^2+\nu^2}\right)^2-r_S(\omega)\right)^2$, $r_S(\omega)\triangleq\frac{\mathrm{E}\left[|S|^2h(S,\bar{\nu},\omega,2);P_{S,\nu}\right]}{\mathrm{E}\left[h(S,\bar{\nu},\omega,0);P_{S,\nu}\right]}$, $\bar{\nu}\triangleq\nu\sigma_{\mathbf{Z}}$, and $h(S,\nu,\omega,k)\triangleq\left(\frac{\omega^2}{\nu^2+\omega^2}\right)^{p+k}\exp\left(-\frac{|S|^2}{\nu^2+\omega^2}\right)$.

By (14) and (15) the empirical estimate of (20) is given by:

$$\hat{P}^{(u)}_e(\omega)=\sum_{k=0}^{1}P_{\boldsymbol{\theta}_k}Q\left(\frac{\tilde{t}^{(u)}_{opt}(\omega)-\tilde{\eta}^{(u)}_{\boldsymbol{\theta}_k}(\omega)}{\sqrt{\tilde{\lambda}^{(u)}_{\boldsymbol{\theta}_k}(\omega)}}\right), \tag{21}$$

where $\tilde{\eta}^{(u)}_{\boldsymbol{\theta}_k}(\omega)\triangleq\frac{\hat{\eta}^{(u)}_{\boldsymbol{\theta}_k}(\omega)}{c(\omega)}=\sum_{n=1}^{N_k}\hat{\varphi}_u\left(\mathbf{X}^{(k)}_n;\omega\right)\xi\left(\mathbf{X}^{(k)}_n;\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right)$, $\tilde{\lambda}^{(u)}_{\boldsymbol{\theta}_k}(\omega)\triangleq\frac{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_k}(\omega)}{c^2(\omega)}=\frac{N_k}{N}\sum_{n=1}^{N_k}\hat{\varphi}^2_u\left(\mathbf{X}^{(k)}_n;\omega\right)\left(\xi\left(\mathbf{X}^{(k)}_n;\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right)-\tilde{\eta}^{(u)}_{\boldsymbol{\theta}_k}\right)^2$, $k=0,1$, $\xi(\mathbf{X};\boldsymbol{\theta}_0,\boldsymbol{\theta}_1)\triangleq\left|\boldsymbol{\theta}_1^H\mathbf{X}\right|^2-\left|\boldsymbol{\theta}_0^H\mathbf{X}\right|^2$, and $\tilde{t}^{(u)}_{opt}(\omega)$ is obtained from (15) by replacing $\hat{\eta}^{(u)}_{\boldsymbol{\theta}_k}$ and $\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_k}$ with $\tilde{\eta}^{(u)}_{\boldsymbol{\theta}_k}(\omega)$ and $\tilde{\lambda}^{(u)}_{\boldsymbol{\theta}_k}(\omega)$.

In the following simulation examples, the vectors $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ were set to $\boldsymbol{\theta}_k\triangleq\frac{1}{\sqrt{p}}\left[1,e^{-i\pi\sin(\vartheta_k)},\ldots,e^{-i\pi(p-1)\sin(\vartheta_k)}\right]^T$, $k=0,1$, with a-priori probabilities of $P_{\boldsymbol{\theta}_0}=0.6$ and $P_{\boldsymbol{\theta}_1}=0.4$ respectively, were $\vartheta_0=0$, $\vartheta_1=\pi/3$ and $p=4$. We consider a BPSK signal with variance σ^2_S and an ϵ-contaminated Gaussian noise model [8] under which the texture component ν in (17) is a binary random variable satisfying $\nu=1$ w.p. $1-\epsilon$ and $\nu=\delta$ w.p. ϵ. The parameters ϵ and δ that control the heaviness of the noise tails were set to 0.2 and 10, respectively. We define the signal-to-noise-ratio (SNR) as $\text{SNR}\triangleq10\log_{10}\sigma^2_S/\sigma^2_{\mathbf{Z}}$. In all examples the sample size was set to $N=300$. The empirical asymptotic probability of error (21) was obtained using two i.i.d. training sequences from $P_{\mathbf{X}|\boldsymbol{\theta}_0}$ and $P_{\mathbf{X}|\boldsymbol{\theta}_1}$ containing $N_0=N_1=3\times10^4$ samples.

In the first example, we compared the asymptotic probability of error (20) to its empirical estimate (21) as a function of ω for SNR$=-8$ [dB]. Observing Fig. 1, one sees that due to the consistency of (21) the compared quantities are very close.

In the second example, we compared the empirical, asymptotic (20) and empirical asymptotic (21) probability of error of the MT-GQLRT to the empirical probability of error of the Bayesian GQLRT and the omniscient Bayesian LRT. The optimal Gaussian MT-function parameter $\omega_{\rm opt}$ was obtained by minimizing (21) over $\Omega=[1,100]$. The empirical probability of error curves were obtained using 10^5 Monte-Carlo simulations. The SNR is used to index the performances as depicted in Fig. 2. One sees that the MT-GQLRT outperforms the non-robust Bayesian GQLRT and, attains classification performance that are significantly closer to those obtained by the Bayesian LRT that, unlike the MT-GQLRT, requires complete knowledge of the conditional likelihood function under each hypothesis.

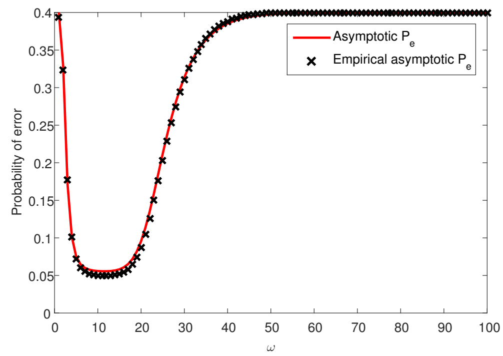

Fig. 1. Asymptotic probability of error (20) and its empirical estimate (21) versus the width parameter ω of the MT-function (18).

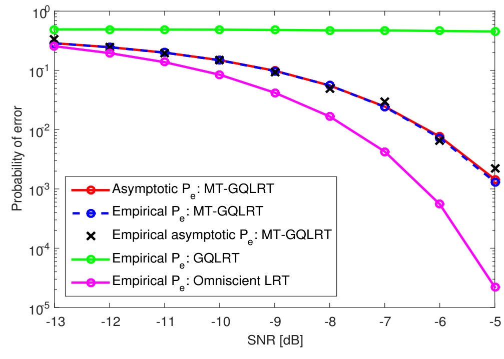

Fig. 2. The empirical, asymptotic (20) and empirical asymptotic (21) probability of error of the MT-GQLRT as compared to the empirical probability of error of the GQLRT and LRT.

5. CONCLUSION

In this paper a new test for Bayesian binary hypothesis testing was developed that is based on a Gaussian LRT after transformation of the conditional probability distribution of the data. By specifying the MT-function in the Gaussian family, the proposed test, called MT-GQLRT, was applied to Bayesian signal classification in non-Gaussian noise. Exploration of other MT-functions may result in additional tests in this class that have different useful properties.

6. REFERENCES

[1] L. L. Scharf, *Statistical signal processing,* Addison-Wesley, 1991.

[2] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis,* Wiley, 1973.

[3] F. J. Jaimes-Romero and D. Munoz-Rodriguez, "Generalized Bayesian hypothesis testing for cell coverage determination," *Vehicular Technology, IEEE Transactions,* vol.49, no.4, pp.1102-1109, Jul 2000.

[4] J. Sohn, N. S. Kim and W. Sung, "A statistical model-based voice activity detection," *Signal Processing Letters, IEEE,* vol.6, no.1, pp.1-3, Jan. 1999

[5] J. Ramírez, J. C. Segura, C. Benítez, L. García and A. Rubio, "Statistical voice activity detection using a multiple observation likelihood ratio test," *IEEE Signal Processing Letters,* vol. 12, no. 10, pp. 689-692, 2005.

[6] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics,* vol. 22, pp. 79-86, 1951.

[7] K. Todros and A. O. Hero, "Measure-transformed quasi likelihood ratio test," *Proceedings of ICASSP 2016*.

[8] E. Ollila, D. E. Tyler, V. Koivunen and H. V. Poor, "Complex elliptically symmetric distributions: survey, new results and applications," *IEEE Transactions on Signal Processing,* vol. 60, no. 1, pp. 5597-5625, 2012.

[9] G. B. Folland, *Real Analysis*. John Wiley and Sons, 1984.

[10] K. Todros and A. O. Hero, "Robust Multiple Signal Classification via Probability Measure Transformation," *IEEE Transactions on Signal Processing,* vol. 63, no. 5, 2015.

[11] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory,* Springer-Verlag, 2006.

[12] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 2011.

[13] P. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals.* p. 39, Cambridge University Press, 2010.

[14] I. S. Dhillon and J. A. Tropp, "Matrix nearness problems with Bregman divergences," *SIAM Journal on Matrix Analysis and Applications,* vol. 29, no. 4, pp. 1120-1146, 2007.