Optimum Joint Detection and Estimation

George V. Moustakides
Department of Electrical and Computer Engineering
University of Patras, 26500 Rion-Patras, GREECE

***Abstract*—We consider the problem of simultaneous binary hypothesis testing and parameter estimation. By defining suitable joint formulations we develop combined detection and estimation strategies that are optimum. Key point of the proposed methodologies constitutes the fact that they integrate both well known approaches, namely Bayesian and Neyman-Pearson.**

I. Introduction

There are important applications in practice where one is confronted with the problem of distinguishing between different hypotheses and, depending on the decision, the need to proceed and estimate a set of relevant parameters. Characteristic examples are: Detection and estimation (segmentation) of objects from images [1]; Retrospective changepoint detection, where one desires to detect a change in statistics but also estimate the time of the change [2]; Defect detection from radiographies, where in addition to detecting presence of defects one would also like to find their position and shape [3]; finally radar applications where one is interested in detecting the presence of a target and also estimate several target characteristics as position, speed, etc. All these applications clearly demand for detection and estimation strategies that address the two subproblems in a *jointly* optimum manner.

In the literature, there are basically two, mainly ad-hoc, approaches that deal with combined problems. The first consists in treating the two subproblems separately and applying in each case the corresponding optimum technique. For instance one can use the Neyman-Pearson optimum test for detection and the optimum Bayesian estimator for parameter estimation to solve the combined problem. As we will see in our analysis, and it is usually the case in combined problems, treating each part separately with the optimum scheme, does not necessarily produce overall optimum performance. The second method consists in using the Generalized Likelihood Ratio Test (GLRT) which detects and estimates at the same time, with the parameter estimation part relying on the maximum likelihood (ML) estimator. Both approaches are not optimum in any finite-sample-size sense.

Surprisingly, one can find *very* limited literature that deals with optimum solutions of joint detection and estimation problems. There are purely Bayesian technique reported in [4], [5] where the performance criterion combines the estimation and detection efficiency in order to capture the collective detection/estimation power. The overall cost is then optimized to yield the optimum combined scheme.

In this work we will consider two different methodologies. The first will resemble the formulation proposed in [4], [5] only here, instead of a purely Bayesian formulation, we adopt a combination of Bayesian and Neyman-Pearson approach. Specifically we will mimic the Neyman-Pearson setup and replace the decision error probabilities, used in classical Neyman-Pearson, with estimation costs. The resulting optimum combined scheme will clearly have similarities with the ones reported in [4], [5]. However, in this part, we will place our main emphasis in proving an interesting optimality property for GLRT and in developing alternative to GLRT detection/estimation structures which rely on estimators that are different from ML.

In the second methodology, we will concentrate on the estimation subproblem and after defining a suitable performance measure for the estimator we will optimize it, assuring in parallel satisfactory performance for the detection subproblem through suitable constraints. This formulation will give rise to novel one- and two-step optimum detection and estimation structures that allow for the trading between detection power and estimation quality. This desirable characteristic is not enjoyed by the previous class of combined schemes.

II. Background

Let us define the problem of interest. Motivated by most applications mentioned in the Introduction, we limit ourselves to the binary hypothesis case with parameters appearing only under the alternative hypothesis. Suppose we are given a finite-sample-size observation signal X for which we have the following two hypotheses

$$\mathsf{H}_0: \ X \sim f_0(X),$$
$$\mathsf{H}_1: \ X \sim f_1(X|\theta), \ \theta \sim \pi(\theta),$$

where $f_0(X), \ f_1(X|\theta)$ and $\pi(\theta)$ are known pdfs. Specifically, we assume that under H_1 the pdf of X contains a collection of random parameters θ for which we have available some prior pdf $\pi(\theta)$, whereas under H_0 the data pdf is completely known. The goal is to develop a mechanism that distinguishes between $\mathsf{H}_0, \mathsf{H}_1$ and, every time this mechanism decides in favor of H_1, it also provides an estimate $\hat{\theta}(X)$ for θ. If D denotes our decision then our combined detection/estimation scheme is comprised of the triplet $\{\delta_0, \delta_1, \hat{\theta}\}$, where $\delta_i(X)$ denotes the probability of a randomized detector for deciding $\mathsf{D} = \mathsf{H}_i$; and $\hat{\theta}(X)$ is a vector function that provides the necessary parameter estimates. Clearly $\delta_i(X) \geq 0$ and $\delta_0(X) + \delta_1(X) = 1$.

Let us recall, very briefly, the basic detection and estimation results when the two subproblems are considered separately.

Neyman-Pearson hypothesis testing: Fix a level $\alpha \in (0,1)$; since D denotes our decision, we are interested in selecting a test (namely the randomization probabilities $\delta_i(X)$) so that the detection probability $\mathsf{P}_1(\mathsf{D} = \mathsf{H}_1)$ is maximized subject to the false alarm constraint $\mathsf{P}_0(\mathsf{D} = \mathsf{H}_1) \leq \alpha$. Equivalently, the previous maximization can be replaced by the minimization of the probability of miss $\mathsf{P}_1(\mathsf{D} = \mathsf{H}_0)$. The optimum detection scheme is the well celebrated likelihood ratio test which takes

978-1-4577-0594-6/11/$26.00 ©2011 IEEE

the following form for our specific setup

$$\frac{f_1(X)}{f_0(X)} = \frac{\int f_1(X|\theta_1)\pi(\theta)\,d\theta}{f_0(X)} \overset{\mathsf{H}_1}{\underset{\mathsf{H}_0}{\gtreqless}} \gamma_{\text{NP}}. \tag{1}$$

In other words we decide H_1 whenever the likelihood ratio exceeds the threshold γ_{NP}; H_0 whenever it falls below and randomize with a probability p when it is equal to the threshold. The threshold γ_{NP} and the probability p are selected to satisfy the false alarm constraint with equality. The randomization probabilities $\delta_0^{\text{NP}}(X), \delta_1^{\text{NP}}(X)$ corresponding to the Neyman-Pearson test are given by

$$\begin{aligned}\delta_0^{\text{NP}}(X) &= \mathbb{1}_{\{\frac{f_1(X)}{f_0(X)}<\gamma_{\text{NP}}\}} + (1-p)\mathbb{1}_{\{\frac{f_1(X)}{f_0(X)}=\gamma_{\text{NP}}\}}\\ \delta_1^{\text{NP}}(X) &= \mathbb{1}_{\{\frac{f_1(X)}{f_0(X)}>\gamma_{\text{NP}}\}} + p\mathbb{1}_{\{\frac{f_1(X)}{f_0(X)}=\gamma_{\text{NP}}\}},\end{aligned} \tag{2}$$

where $\mathbb{1}_{\mathcal{A}}$ denotes the indicator function of the set $\mathcal{A}$.

Bayesian parameter estimation: Suppose that we know *with certainty* that the observations X come from hypothesis H_1, then we are interested in providing an estimate $\hat{\theta}(X)$ for the parameters θ. We measure the quality of our estimate with the help of a cost function $C(\hat{\theta},\theta) \geq 0$. We would like to select the optimum estimator in order to minimize the average cost $\mathsf{E}_1[C(\hat{\theta}(X),\theta)]$, where expectation is with respect to X, θ.

From [6, Page 142] we have that the optimum Bayesian estimator is the following minimizer (provided it exists)

$$\hat{\theta}_o(X) = \arg\inf_U \mathcal{C}(U|X), \tag{3}$$

where $\mathcal{C}(U|X)$ is the posterior cost function

$$\mathcal{C}(U|X) = \mathsf{E}_1[C(U,\theta)|X] = \frac{\int C(U,\theta) f_1(X|\theta)\pi(\theta)\,d\theta}{\int f_1(X|\theta)\pi(\theta)\,d\theta} \tag{4}$$

and expectation, as we can see from the last equality, is with respect to θ for given X. Finally we denote the optimum posterior cost as $\mathcal{C}_o(X)$, that is,

$$\mathcal{C}_o(X) = \inf_U \mathcal{C}(U|X) = \mathcal{C}(\hat{\theta}_o(X)|X). \tag{5}$$

This quantity will play a significant role in the development of our theory as it constitutes a genuine quality index for the estimate $\hat{\theta}_o(X)$.

Let us now consider the combined problem. We recall that the hypothesis testing part distinguishes between H_0 and H_1. As we have seen, the Neyman-Pearson approach provides the best possible detection structure for controlling and optimizing the corresponding decision error probabilities. However with a decision mechanism that focuses on the decision errors, we cannot necessarily guarantee efficiency for the estimation part. Consequently, we understand, that the detection part cannot be treated independently from estimation. Following this rationale, in the next two sections we propose two possible approaches for the joint problem.

III. Bayesian/Neyman-Pearson-like Formulation

Consider the case where the true hypothesis is H_1 and the true parameter vector is θ. We then distinguish two possible costs: $C(\hat{\theta},\theta) \geq 0$ denotes the cost for deciding $\mathsf{D} = \mathsf{H}_1$ and providing the estimate $\hat{\theta}(X)$, while $D(\theta) \geq 0$ the cost for deciding in favor of H_0 and therefore providing no estimate. We can now define the average cost under H_1 as follows

$$\mathcal{J}(\delta_0,\delta_1,\hat{\theta}) = \mathsf{E}_1[C(\hat{\theta}(X),\theta)\delta_1(X) + D(\theta)\delta_0(X)], \tag{6}$$

where expectation is with respect to X, θ. We observe that the average cost incorporates the complete detection and estimation structure, that is, the detector (expressed through $\delta_i(X)$) and the estimator (expressed through $\hat{\theta}(X)$). We should note that the second part involving the cost $D(\theta)$ measures the contribution of the missing detections in the overall cost.

We can now mimic the Neyman-Pearson formulation and instead of simply minimizing the probability of miss which is the usual practice in the Neyman-Pearson approach we can replace $\mathsf{P}_1(\mathsf{D} = \mathsf{H}_0)$ with $\mathcal{J}(\delta_0,\delta_1,\hat{\theta})$. In other words we ask to *minimize the average cost* under H_1. This minimization will be performed subject to the familiar false alarm constraint $\mathsf{P}_0(\mathsf{D} = \mathsf{H}_1) \leq \alpha$. The constrained optimization problem and the corresponding optimum solution are presented in the next theorem.

Theorem 1. *The optimum combined detection and estimation scheme that solves the constrained optimization problem*

$$\inf_{\delta_0,\delta_1,\hat{\theta}} \mathcal{J}(\delta_0,\delta_1,\hat{\theta}); \text{ subject to } \mathsf{P}_0(\mathsf{D} = \mathsf{H}_1) \leq \alpha$$

is given by the Bayesian estimator $\hat{\theta}_o(X)$ defined in (3) *and the following detector*

$$\mathcal{D}(X) \overset{\mathsf{H}_1}{\underset{\mathsf{H}_0}{\gtreqless}} \mathcal{C}_o(X), \text{ if } \mathsf{P}_0(\mathcal{D}(X) \geq \mathcal{C}_o(X)) \leq \alpha, \tag{7}$$

$$\frac{f_1(X)}{f_0(X)}[\mathcal{D}(X) - \mathcal{C}_o(X)] \overset{\mathsf{H}_1}{\underset{\mathsf{H}_0}{\gtreqless}} \gamma, \text{ if } \mathsf{P}_0(\mathcal{D}(X) \geq \mathcal{C}_o(X)) > \alpha, \tag{8}$$

where $\mathcal{D}(X) = \mathsf{E}_1[D(\theta)|X]$; $\mathcal{C}_o(X)$ is defined in (5)*; $f_1(X) = \int f_1(X|\theta)\pi(\theta)d\theta$; and the threshold γ is selected so that the test in* (8) *satisfies the false alarm constraint with equality.*

Proof: First consider the case $\mathsf{P}_0(\mathcal{D}(X) \geq \mathcal{C}_o(X)) \leq \alpha$, then for the average cost we can write

$$\begin{aligned}\mathcal{J}(\delta_0,\delta_1,\hat{\theta}) &= \mathsf{E}_1[C(\hat{\theta}(X),\theta)\delta_1(X) + D(\theta)\delta_0(X)]\\ &= \int[\mathcal{C}(\hat{\theta}(X)|X)\delta_1(X) + \mathcal{D}(X)\delta_0(X)]f_1(X)dX\\ &\geq \int\{\mathcal{C}_o(X)f_1(X)\delta_1(X) + \mathcal{D}(X)f_1(X)\delta_0(X)\}dX\\ &\geq \int \min\{\mathcal{C}_o(X),\mathcal{D}(X)\}f_1(X)dX\\ &= \mathsf{E}_1[C(\hat{\theta}_o(X),\theta)\mathbb{1}_{\{\mathcal{D}(X)\geq\mathcal{C}_o(X)\}} + D(\theta)\mathbb{1}_{\{\mathcal{D}(X)<\mathcal{C}_o(X)\}}]\\ &= \mathcal{J}(\delta_0^o,\delta_1^o,\hat{\theta}_o),\end{aligned}$$

where $\delta_0^o(X) = \mathbb{1}_{\{\mathcal{D}(X)<\mathcal{C}_o(X)\}}$, $\delta_1^o(X) = \mathbb{1}_{\{\mathcal{D}(X)\geq\mathcal{C}_o(X)\}}$, i.e. the test in (7) expressed in terms of randomization probabilities. Also for simplicity in our proof we assumed $\mathsf{P}_1(\mathcal{D}(X) = \mathcal{C}_o(X)) = 0$. If this is not true then a randomization is required every time $\mathcal{D}(X) = \mathcal{C}_o(X)$. Note that the first inequality in the previous relations is true because $\mathcal{C}(\hat{\theta}(X)|X) \geq \inf_U \mathcal{C}(U|X) = \mathcal{C}_o(X)$ and from (5) we have equality iff $\hat{\theta}(X) = \hat{\theta}_o(X)$. Furthermore the last inequality is true because $\delta_i(X) \geq 0$ and $\delta_0(X) + \delta_1(X) = 1$. Equality in this last inequality is assured iff $\delta_i(X) = \delta_i^o(X)$, that is, with the randomization probabilities corresponding to the detector defined in (7). The triplet $\{\delta_0^o,\delta_1^o,\hat{\theta}_o\}$ is thereby optimum since it also satisfies the false alarm constraint due to our initial assumption that $\mathsf{P}_0(\mathcal{D}(X) \geq \mathcal{C}_o(X)) \leq \alpha$.

Consider now the case where $\mathsf{P}_0(\mathcal{D}(X) \geq \mathcal{C}_o(X)) > \alpha$ and assume for simplicity that $\mathcal{G}(X) = \frac{f_1(X)}{f_0(X)}[\mathcal{D}(X) - \mathcal{C}_o(X)]$, when seen as a random variable due to the randomness of X, has no atoms, namely, no single value attained by $\mathcal{G}(X)$ has nonzero probability. We will first show that there exists $\gamma > 0$ such that $\mathsf{P}_0(\mathcal{G}(X) \geq \gamma) = \alpha$. In other words that there is threshold γ so that the test in (8) satisfies the false alarm constraint with equality. Define the difference

$$\phi(\gamma) = \mathsf{P}_0\left(\frac{f_1(X)}{f_0(X)}[\mathcal{D}(X) - \mathcal{C}_o(X)] \geq \gamma\right) - \alpha.$$

Then we observe that, due to the assumption $\mathsf{P}_0(\mathcal{D}(X) \geq \mathcal{C}_o(X)) > \alpha$, we have $\phi(0) > 0$. Also $\lim_{\gamma\to\infty}\phi(\gamma) = -\alpha < 0$. Consequently there exists $\gamma > 0$ where we have $\phi(\gamma) = 0$, namely, the test in (8) satisfies the false alarm constraint with equality. Let us now prove that this test minimizes the average cost among all tests that satisfy the false alarm constraint.

If $\gamma > 0$ is the quantity defined above, then for any combined test satisfying the false alarm constraint we have

$$\begin{aligned}
&\mathcal{J}(\delta_0,\delta_1,\hat{\theta}) + \gamma\alpha \geq \mathcal{J}(\delta_0,\delta_1,\hat{\theta}) + \gamma\mathsf{P}_0(\mathsf{D}=\mathsf{H}_1)\\
&\quad= \mathsf{E}_1[C(\hat{\theta}(X),\theta)\delta_1(X) + D(\theta)\delta_0(X)] + \gamma\mathsf{E}_0[\delta_1(X)]\\
&\quad= \textstyle\int[\mathcal{C}(\hat{\theta}(X)|X)\delta_1(X) + \mathcal{D}(X)\delta_0(X)]f_1(X)dX\\
&\qquad\qquad + \textstyle\int\gamma\delta_1(X)f_0(X)dX\\
&\geq \textstyle\int\{[\mathcal{C}_o(X)f_1(X)+\gamma f_0(X)]\delta_1(X)+\mathcal{D}(X)f_1(X)\delta_0(X)\}dX\\
&\quad\geq \textstyle\int\min\{[\mathcal{C}_o(X)f_1(X)+\gamma f_0(X)],\mathcal{D}(X)f_1(X)\}dX\\
&\quad= \mathsf{E}_1[C(\hat{\theta}_o(X),\theta)\mathbb{1}_{\{\mathcal{G}(X)\geq\gamma\}} + D(\theta)\mathbb{1}_{\{\mathcal{G}(X)<\gamma\}}]\\
&\qquad\qquad + \gamma\mathsf{P}_0(\mathcal{G}(X)\geq\gamma) = \mathcal{J}(\delta_0^o,\delta_1^o,\hat{\theta}_o) + \gamma\alpha.
\end{aligned}$$

Comparing the first with last term, we conclude that $\mathcal{J}(\delta_0,\delta_1,\hat{\theta}) \geq \mathcal{J}(\delta_0^o,\delta_1^o,\hat{\theta}_o)$ where $\hat{\theta}_o(X)$ is defined in (3) and $\delta_0^o(X) = \mathbb{1}_{\{\mathcal{G}(X)<\gamma\}}$, $\delta_1^o(X) = \mathbb{1}_{\{\mathcal{G}(X)\geq\gamma\}}$, namely the test in (8) expressed with randomization probabilities. ∎

As we can see from both versions of the test in (7),(8) $\mathcal{C}_o(X)$ plays the role of a quality index for the estimate $\hat{\theta}_o(X)$. Indeed the larger this quantity, the less reliable the estimate is considered, since the less chance the test has to decide in favor of H_1. We will see that $\mathcal{C}_o(X)$ continues to enjoy the same role in all test that we introduce in the sequel.

We would like to point out that the proof and the results developed so far, present definite similarities with [4]. As far as Theorem 1 is concerned the pure Bayesian approach in [4] captures only the test in (8) and not the version in (7). Even though this difference is not dramatic we continue in this direction because we would like to present an interesting optimality result for GLRT not mentioned in [4] and, furthermore, with the help of Theorem 1 introduce alternative to GLRT tests. Let us first apply Theorem 1 to demonstrate an important optimality property for GLRT.

A. Finite sample size optimality of GLRT

Suppose that under H_1, parameter θ can assume a finite number of possible values $\theta \in \{\theta_1,\ldots,\theta_N\}$ with prior probabilities $\pi_1,\ldots,\pi_N$. Consider now the following cost functions $C(\hat{\theta},\theta) = \mathbb{1}_{\{\hat{\theta}\neq\theta\}}$ and $D(\theta) = 1$. It is then very easy to realize that $\mathcal{J}(\delta_0,\delta_1,\hat{\theta})$ expresses *the probability of making an error under* H_1. Indeed, from (6) we can see that the first part involving the cost $C(\hat{\theta},\theta)$ corresponds to the case where we select correctly the hypothesis but we make an error in the selection of the parameter, whereas the second term involving the cost $D(\theta)$ is the probability to select incorrectly the hypothesis. Consequently the resulting average cost is a legitimate performance criterion that makes a lot of sense.

We would like to minimize the error probability under H_1 assuring, at the same time, that the false alarm probability is no larger than α. We should point out that if in Theorem 1 we select the prior pdf $\pi(\theta)$ to be a collection of Dirac function, then we can easily accommodate the case where θ assumes a finite number of values. Note that $f_1(X) = \sum_{k=1}^N f_1(X|\theta_k)\pi_k$ and for $U \in \{\theta_1,\ldots,\theta_N\}$, we can write

$$\mathcal{C}(U|X) = 1 - \frac{\mathbb{1}_{\{U=\theta_n\}}f_1(X|\theta_n)\pi_n}{f_1(X)}$$

$$\mathcal{C}_o(X) = 1 - \frac{\max_{\theta_n} f_1(X|\theta_n)\pi_n}{f_1(X)}$$

$$\hat{\theta}_o(X) = \arg\max_{\theta_n} f_1(X|\theta_n)\pi_n.$$

As expected, $\hat{\theta}_o(X)$ is simply the MAP estimator. Applying now Theorem 1 and observing that $\mathcal{D}(X) = 1$ we first conclude that (7) leads to a trivial test. Consequently the optimum test that minimizes the error probability under H_1 is the version in (8) which takes the form

$$\frac{\max_{\theta_n} f_1(X|\theta_n)\pi_n}{f_0(X)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \gamma.$$

Additionally, if the prior is uniform, that is, $\pi_n = 1/N$, then the previous test is equivalent to

$$\frac{\max_{\theta_n} f_1(X|\theta_n)}{f_0(X)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} N\gamma = \gamma'$$

which is the GLRT while $\hat{\theta}_o(X) = \arg\max_{\theta_n} f_1(X|\theta_n)$ is the ML estimator. We thus conclude that by adopting as cost function the one that leads to the MAP estimator and then assuming uniform prior for the parameter θ, we can prove finite-sample-size optimality for GLRT. This interesting result in not reported in [4] and, furthermore, GLRT is not known to enjoy any finite-sample-size optimality property.

Similar conclusion can be drawn when θ assumes a continuum of values and we use as cost function $C(\hat{\theta},\theta) = \mathbb{1}_{\{|\hat{\theta}-\theta|>\varDelta\}}$, where $0 < \varDelta \ll 1$, i.e. the one that leads to the MAP estimator in the classical Bayesian estimation theory [6, Pages 145-147]. Considering also $D(\theta) = 1$ and uniform prior for θ, if we follow similar steps as in the finite case, we can show that the resulting test takes the familiar GLRT form

$$\frac{\sup_\theta f_1(X|\theta)}{f_0(X)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \gamma, \tag{9}$$

with the corresponding optimum estimator being equal to the ML estimator $\hat{\theta}_o(X) = \arg\sup_\theta f_1(X|\theta)$.

B. Alternative tests

If one is not content with ML and desires to use instead MMSE or minimum mean absolute error (MMAE) estimates, the question is whether it is possible to develop tests, that employ these estimates, and can replace GLRT. Theorem 1 gives the general framework that allows for the development of such results. We recall that key assumption in obtaining the GLRT is that the prior $\pi(\theta)$ is uniform; assumption that we also adopt for the MMSE and MMAE criterion.

1) MMSE Detection/Estimation: We consider the cost function $C(\hat{\theta},\theta)=\|\hat{\theta}-\theta\|^2$. The second cost function $D(\theta)$ will be specified in the sequel. From Theorem 1 we have that the estimator in the optimum combined scheme is the minimizer of the conditional mean square error which is the conditional mean of θ given the data X [6, Page 153]. Under uniform prior we can thus write

$$\hat{\theta}_o(X)=\mathsf{E}_1[\theta|X]=\frac{\int\theta f_1(X|\theta)d\theta}{\int f_1(X|\theta)d\theta}.$$

Similarly for the optimum posterior cost $\mathcal{C}_o(X)$ we have

$$\mathcal{C}_o(X)=\frac{\int\|\theta\|^2 f_1(X|\theta)d\theta}{\int f_1(X|\theta)d\theta}-\|\hat{\theta}_o(X)\|^2.$$

Let us now apply the optimum tests proposed in Theorem 1. Recalling that

$$\mathcal{D}(X)=\mathsf{E}_1[D(\theta)|X]=\frac{\int D(\theta)f_1(X|\theta)d\theta}{\int f_1(X|\theta)d\theta},$$

we have that the first version in (7) is equivalent to

$$\|\hat{\theta}_o(X)\|^2 \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \frac{\int[\|\theta\|^2-D(\theta)]f_1(X|\theta)d\theta}{\int f_1(X|\theta)d\theta}$$

and the second in (8) equivalent to

$$\frac{\int f_1(X|\theta)d\theta}{f_0(X)}\left\{\|\hat{\theta}_o(X)\|^2-\frac{\int[\|\theta\|^2-D(\theta)]f_1(X|\theta)d\theta}{\int f_1(X|\theta)d\theta}\right\}\underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}}\gamma.$$

For $D(\theta)$ it seems reasonable to select $D(\theta)=\|\theta\|^2$. This becames particularly apparent when $f_0(X)=f_1(X|\theta=0)$, that is, when we are interested in detecting whether the value of a parameter has changed from the nominal value $\theta=0$ to some alternative that we also like to estimate. Consequently, when under H_1 we decide $\mathsf{D}=\mathsf{H}_0$, it is as if we select $\theta=0$ which yields a squared error $\|\theta-0\|^2$. With this selection of the cost function $D(\theta)$, the first version of the test is never used because it always decides in favor of H_1. For the second version we have the following simple form

$$\frac{\int f_1(X|\theta)d\theta}{f_0(X)}\|\hat{\theta}_o(X)\|^2\underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}}\gamma.$$

As we can see, the estimate plays an explicit role in the decision process. This is the equivalent of the GLRT test defined in (9) but with the MMSE replacing the ML estimation.

2) MMAE Detection/Estimation: Following similar steps for the case $C(\hat{\theta},\theta)=|\hat{\theta}-\theta|$ and assuming for simplicity that $\hat{\theta},\theta$ are scalars, we end up with the optimum estimator being the conditional median [6, Page 153], that is, the solution of the equation

$$\hat{\theta}_o(X)=\arg\left\{\hat{\theta}:\frac{\int_{-\infty}^{\hat{\theta}}f_1(X|\theta)d\theta}{\int f_1(X|\theta)d\theta}=0.5\right\}. \tag{10}$$

Furthermore if we select $D(\theta)=|\theta|$ which, again, makes sense under the frame we discussed in the MMSE criterion, we end up with a unique version for the optimum test which has the following interesting form

$$\frac{\int_0^{\hat{\theta}_o}\theta f_1(X|\theta)d\theta}{f_0(X)}\underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}}\gamma.$$

This test is the alternative of the GLRT when we adopt the MMAE in place of the ML estimator.

IV. Formulation with Constrained Detection

One may argue that the Bayesian/Neyman-Pearson-like methodology of the previous section, presents an important weakness due to the necessity to specify the cost induced by the missed detections. As we realize from the previous examples, proposing a suitable $D(\theta)$ is significantly more arbitrary than specifying the cost $C(\hat{\theta},\theta)$.

In this section we follow a different direction that completely bypasses the need to define this additional function. As we will see, depending on how we incorporate the notion of *reliable estimate* into our setup, this new formulation will yield one- and two-step detection/estimation strategies that are completely novel.

A. One-step tests

As before, the combined scheme consists of the triplet $\{\delta_0,\delta_1,\hat{\theta}\}$. We like to decide between $\mathsf{H}_0,\mathsf{H}_1$ but in the following sense: We decide in favor of H_1 *only if our decision can lead to a reliable estimate of* θ. In other words we do not care to distinguish the null hypothesis from the alternative if the latter cannot provide reliable estimates. There are various application where this reasoning makes sense. For example in Image Processing when one is interested in segmenting objects in images, detection of an object is useless unless we can find its border (estimation).

Focusing on the estimation subproblem, we can now define the following performance measure

$$\mathcal{J}(\delta_0,\delta_1,\hat{\theta})=\mathsf{E}_1[C(\hat{\theta},\theta)|\mathsf{D}=\mathsf{H}_1]. \tag{11}$$

We use the classical Bayesian cost function for the estimator $\hat{\theta}$ but we condition on the event that we have an estimate only when we decide in favor of H_1. As before the proposed performance measure depends on the complete detection/estimation structure. Clearly the goal is to minimize $\mathcal{J}(\delta_0,\delta_1,\hat{\theta})$ however, we observe that this criterion assesses only the quality of the estimation subproblem. In order to guarantee satisfactory performance for detection we will impose suitable constraints.

We propose to control the two decision error probabilities as follows: $\mathsf{P}_0(\mathsf{D}=\mathsf{H}_1)\le\alpha$ which is our familiar false alarm constraint, but also, $\mathsf{P}_1(\mathsf{D}=\mathsf{H}_0)\le\beta$ which is a constraint on the probability of miss. We need to select $1>\beta\ge\beta_{\rm NP}(\alpha)$ where $\beta_{\rm NP}(\alpha)$ is the probability of miss of the Neyman-Pearson test defined in (1), because no test can have smaller probability of miss than the Neyman-Pearson test. What we are actually proposing here is *to sacrifice part of the detection power* as compared to the Neyman-Pearson test (by allowing more misses), in the hope that this will induce a significant improvement in the estimation quality. We have the following theorem that solves the constrained optimization problem.

Theorem 2. *Consider the two constraints* $\mathsf{P}_0(\mathsf{D}=\mathsf{H}_1)\le\alpha$ *and* $\mathsf{P}_1(\mathsf{D}=\mathsf{H}_0)\le\beta$*, where* $1>\alpha>0$ *and* $1>\beta\ge\beta_{\rm NP}(\alpha)$ *with* $\beta_{\rm NP}(\alpha)$ *denoting the probability of miss of the Neyman-Pearson test. Let* $\lambda_o>0$ *be the solution of the equation*

$$\mathsf{P}_1\left(\lambda_o\ge\mathcal{C}_o(X)\right)=1-\beta,$$

where $\mathcal{C}_o(X)$ *is defined in* (5)*. Then the optimum combined scheme that minimizes* $\mathcal{J}(\delta_0,\delta_1,\hat{\theta})$ *in* (11) *under the two error constraints is comprised of the Bayesian estimator* $\hat{\theta}_o(X)$ *defined in* (3) *and the following two versions for the decision* *rule*

$$\mathcal{C}_o(X) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\lesseqgtr}} \lambda_o, \quad \text{if } \alpha \geq \mathsf{P}_0\left(\lambda_o \geq \mathcal{C}_o(X)\right) \tag{12}$$

$$\frac{f_1(X)}{f_0(X)}[\lambda - \mathcal{C}_o(X)] \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \gamma, \quad \text{if } \alpha < \mathsf{P}_0\left(\lambda_o \geq \mathcal{C}_o(X)\right), \tag{13}$$

where in (13) λ, γ *are selected so that the two error probability constraints are satisfied with equality.*

Proof: Although the results appear to be similar, the proof of this theorem is more involved than the one of Theorem 1. Details can be found in [7]. ■

B. Two-step tests

In most applications it is undesirable to give up part of the detection power. This is for example the case in radars where it is still helpful to detect a target even if we cannot reliably estimate its parameters. For such problems we propose the following two-step approach: We use an initial detector to distinguish between H_0 and H_1; whenever this test decides in favor of H_1 then, at a second step, we compute the estimate $\hat{\theta}(X)$ and *employ a second test that decides whether the estimate is reliable or unreliable*, denoted as $\mathsf{H}_{1r}, \mathsf{H}_{1u}$ respectively. Consequently we make *three* decisions $\mathsf{H}_0, \mathsf{H}_{1r}, \mathsf{H}_{1u}$ with the union of the last two corresponding to hypothesis H_1. As we can see, we "trust" the estimate $\hat{\theta}(X)$ only when we decide in favor of H_{1r}, but we still have detection even if we discard the estimate as unreliable.

For the first test we use our familiar randomization probabilities $\{\delta_0(X), \delta_1(X)\}$ and for the second we employ a new pair $\{q_{1r}(X), q_{1u}(X)\}$. The latter functions are the randomization probabilities needed to decide between reliable/unreliable estimation *given* that the first test decided in favor of H_1. Therefore we have $q_{1r}(X), q_{1u}(X) \geq 0$ and $q_{1r}(X) + q_{1u}(X) = 1$. We conclude that the complete set of quantities needed to be specified is now $\{\delta_0, \delta_1, q_{1r}, q_{1u}, \hat{\theta}\}$.

For the first test we minimize the probability of miss subject to the false alarm constraint $\mathsf{P}_0(\mathsf{D} = \mathsf{H}_1) \leq \alpha$. This of course yields as optimum the Neyman-Pearson test defined in (1). Having identified the test in the first step we proceed to the second which involves the estimator $\hat{\theta}(X)$ and the second decision mechanism that uses $q_{1r}(X), q_{1u}(X)$ to label the estimate as reliable/unreliable. As in the previous subsection we define the conditional cost for the estimator

$$\mathcal{J}(\delta_0^{\mathrm{NP}}, \delta_1^{\mathrm{NP}}, q_{1r}, q_{1u}, \hat{\theta}) = \mathsf{E}_1[C(\hat{\theta}(X), \theta) | \mathsf{D} = \mathsf{H}_{1r}], \tag{14}$$

with $\delta_i^{\mathrm{NP}}(X)$ defined in (2) and expressing the fact that for the first test we use the Neyman-Pearson test. Conditioning is now with respect to the event $\{\mathsf{D} = \mathsf{H}_{1r}\}$ since this is the only case where the estimate $\hat{\theta}(X)$ is accepted.

We recall that not all detections lead to reliable estimates. Since we do not want to characterise a great deal of our detections as providing unreliable estimates, we need to impose the constraint $1 - \beta \leq \mathsf{P}_1(\mathsf{D} = \mathsf{H}_{1r})$ in order to control their probability. In other words the probability of reliable estimates must be no smaller than a prescribed level $1 - \beta$ which, of course, cannot exceed the level of the original detections of the first step which is $1 - \beta_{\mathrm{NP}}(\alpha)$; consequently β must satisfy $1 > \beta \geq \beta_{\mathrm{NP}}(\alpha)$. We have the following theorem that identifies the estimator and the second test that optimize the cost in (14).

Theorem 3. *Let* $1 > \beta \geq \beta_{\mathrm{NP}}(\alpha)$*, then the estimator and the test that minimize the average conditional cost defined in* (14) *subject to the constraint* $1 - \beta \leq \mathsf{P}_1(\mathsf{D} = \mathsf{H}_{1r})$ *are the Bayesian estimator* $\hat{\theta}_o(X)$ *defined in* (3) *and for the test we have*

$$\mathcal{C}_o(X) \underset{\mathsf{H}_{1u}}{\overset{\mathsf{H}_{1r}}{\lesseqgtr}} \lambda, \tag{15}$$

where λ *is selected to satisfy the constraint with equality.*

Proof: The proof presents no particular difficulties. Details can be found in [7]. ■

In all three detection/estimation schemes appearing in the corresponding three theorems, the quantity $\mathcal{C}_o(X)$ plays the role of a quality index for the estimate $\hat{\theta}_o(X)$. This is particularly apparent in the last case where $\mathcal{C}_o(X)$ is explicitly used to label the corresponding estimate as reliable/unreliable by simply comparing this quantity to a threshold. Finally in the last two combined schemes we observe that there is the possibility to trade, in a very controlled way, detection performance with estimation efficiency. This useful characteristic is not enjoyed by the first schemes presented in Section III, including the popular GLRT.

Numerical examples that demonstrate the possibility of enjoying significant performance gains using the latter schemes over the conventional methodology of treating the two subproblems separately, can be found in [7].

V. Conclusion

We have presented two possible formulations of the joint detection and estimation problem and developed the corresponding optimum solutions. The first formulation combines the Bayesian method for estimation with the Neyman-Pearson approach for detection. This leads to a number of interesting combined schemes that can be used as alternatives to GLRT and proves an interesting finite-sample-size optimality property for this test. In the second formulation we propose a Bayesian approach for the estimation subproblem assuring in parallel the satisfactory performance of the detection part through suitable constraints. This results in one- and two-step optimum schemes that can trade detection power for estimation quality under a completely controlled frame.

References

[1] B.-N. Vo et al., "Joint detection and estimation of multiple objects from image observations," *IEEE Trans. Sig. Proc.*, vol. 58, no. 10, pp. 5129-5141, Oct. 2010.

[2] S. Boutoille et al., "A hybrid fusion system applied to off-line detection and change-points estimation," *Inf. Fusion*, vol. 11, no. 4, pp. 325-337, Oct. 2010.

[3] L. Fillatre et al., "A simple algorithm for defect detection from a few radiographies," *J. Comp.*, vol. 2, no. 6, pp. 26-34, Aug. 2007.

[4] D. Middleton, R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inf. Th.*, vol. IT-14, no. 3, pp. 434-444, May 1968.

[5] M. Rezaeian, B.-N. Vo, "Error bounds for joint detection and estimation of a single obsject with random finite set observation," *IEEE Trans. Sig. Proc.*, vol. 58, no. 3, pp. 1493-1506, March 2010.

[6] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed., Springer, 1998.

[7] G. V. Moustakides et al., "Joint detection and estimation: optimum tests and applications", submitted to the *IEEE Trans. Inf. Th.*, available at `http://arxiv.org/pdf/1101.5084`.