

Finite Sample Size Optimality of GLR Tests

George V. Moustakides, *Senior Member, IEEE*

Abstract

In several interesting applications one is faced with the problem of simultaneous binary hypothesis testing and parameter estimation. Although such joint problems are not infrequent, there exist no systematic analysis in the literature that treats them effectively. Existing approaches consider the detection and the estimation subproblems separately, applying in each case the corresponding optimum strategy. As it turns out the overall scheme is not necessarily optimum since the criteria used for the two parts are usually incompatible. In this article we propose a mathematical setup that considers the two problems jointly. Specifically we propose a meaningful combination of the Neyman-Pearson and the Bayesian criterion and we provide the optimum solution for the joint problem. In the resulting optimum scheme the two parts interact with each other, producing detection/estimation structures that are completely novel. Notable side-product of our work is the proof that the well known GLR test is finite-sample-size optimum under this combined sense.

Index Terms

GLRT, Joint detection/estimation.

I. INTRODUCTION

There exist applications in practice where one must resolve the following problem: decide between two hypotheses H_0 and H_1 and then, depending on the decision, estimate a corresponding set of parameters θ_0 or θ_1 . Characteristic example of a problem that can be formulated under this combined detection/estimation framework is *target detection and localization* by MIMO radar, where one is not only interested in the classical radar detection problem (presence/absence of a target) but also in estimating its position every time a target is declared present [1], [2]. A second example is *retrospective changepoint*

G.V. Moustakides is with the Department of Electrical and Computer Engineering, University of Patras, 26500 Rion, Greece, e-mail: moustaki@upatras.gr.

Manuscript received , 2009; revised , 2009.

detection where we are interested in determining whether there is a point in our samples after which the statistical behavior of the data has changed and, once it is detected then localize this point of interest [3], [4]. Clearly segmentation problems can be formulated as retrospective changepoint detection problems.

We would like to emphasize that our goal is not to solve the pure detection problem in the presence of unknown parameters (for this case the parameter estimation subproblem constitutes only an auxiliary step). In our approach the estimation part is a *vital goal* in the whole setup and of the same importance as the detection part. This is clearly the case in the two examples we mentioned before, where the localization of the target in the first and of the changepoint in the second, are of the same importance as the detection part. Current literature does not treat combined problems systematically and the aim of this article is to cover exactly this gap.

Before introducing in a formal way the combined problem, let us first recall, briefly, the corresponding formulation and the available finite-sample-size optimality results for detection and parameter estimation. For both problems we assume the existence of a random data vector $\mathcal{X} \in \mathbb{R}^N$ of length N .

Binary hypothesis testing: We consider the following two hypotheses H_0, H_1 for \mathcal{X}

$$H_i : \mathcal{X} \sim f_i(X|\theta_i), \quad i = 0, 1, \quad (1)$$

where “ \sim ” means “distributed according to” and $f_i(X|\theta_i)$, $i = 0, 1$, are two distinct pdfs with θ_i denoting a vector of parameters under each hypothesis. Given a realization X of \mathcal{X} , one must decide between the two hypotheses H_0 and H_1 . If $d \in \{0, 1\}$ denotes our decision, then under a Neyman-Pearson formulation we are interested in the following constrained minimization problem

$$\min P(d = 0|H_1), \quad \text{subject to } P(d = 1|H_0) \leq \alpha, \quad (2)$$

where $P(\cdot)$ denotes probability and $\alpha \in (0, 1)$ the maximal allowable false alarm level. Optimization is performed over all decision strategies that satisfy the constraint.

Under a *finite-sample-size* setting, when the two pdfs are completely known, i.e. there are no unknown parameters, the optimum test is the celebrated Likelihood Ratio test. If the pdfs have *unknown* parameters, except the very rare case where a uniformly most powerful test can be found, the problem in (2) is not well defined and one needs to resort to min-max formulations for which no systematic solution exists. In this case it is very common to use the Generalized Likelihood Ratio (GLR) test

$$\frac{\sup_{\theta_1 \in \Theta_1} f_1(X|\theta_1)}{\sup_{\theta_0 \in \Theta_0} f_0(X|\theta_0)} \underset{H_0}{\overset{H_1}{\geq}} \lambda, \quad (3)$$

where Θ_i denotes some a-priori known set of values for θ_i . For the GLR test there is no finite-sample-size optimality result. In fact there are counterexamples against this claim [5], [6], [7]. Nevertheless the use of the GLR test is widespread in applications and one important sideproduct of our analysis is the demonstration that this popular detection scheme is in fact finite-sample-size optimum *under the combined detection and estimation formulation* we are proposing here. We would like to stress that this is no direct contradiction with the counterexamples reported in [5], [6], [7] since in these references the GLR test is evaluated as a pure detector and not in the combined sense we are proposing in this article.

Regarding the problem in (2), if we assume that the parameters θ_i are *random* with known prior pdfs $\pi_i(\theta_i), i = 0, 1$, then again (2) has a well defined solution which is the likelihood ratio test between the two marginal pdfs $f_i(X) = \int f_i(X|\theta_i)\pi_i(\theta_i)d\theta_i$.

Parameter Estimation: In this problem, we assume that \mathcal{X} has a pdf $f(X|\theta)$ where θ , as before, denotes a vector of parameters. If X is a realization of \mathcal{X} , the goal is to use the data X in order to provide an estimate $\hat{\theta}$ for θ . Under a finite-sample-size setup, optimum estimation structures are available for the Bayesian formulation and only when θ is assumed to be random with a known prior pdf $\pi(\theta)$. Specifically, if $C(\hat{\theta}, \theta)$ denotes the cost of providing the estimate $\hat{\theta}$ when the true parameter value is θ , then the optimum estimator that *minimizes the average cost* is

$$\hat{\theta} = \arg \inf_U \int C(U, \theta) f(X|\theta) \pi(\theta) d\theta. \quad (4)$$

With proper choice of the cost function $C(\hat{\theta}, \theta)$, this formula gives rise to a number of well known estimators as the MAP, the conditional mean or the conditional median.

Next we will combine the two problems and after defining a meaningful performance measure we will develop the optimum detection/estimation structure for the joint problem.

II. COMBINED DETECTION AND ESTIMATION

As we realize from the previous discussion, in both problems, finite-sample-size optimum solutions exist only if we assume that the parameters are random with some known prior. It is therefore natural to expect that the same assumption will be transferred to the more general combined problem. With this observation in mind, let us define the problem of interest.

Consider a random data vector $\mathcal{X} \in \mathbb{R}^N$ and the following two hypotheses H_0, H_1 :

$$H_i : \mathcal{X} \sim f_i(X|\theta_i) \text{ with prior pdf } \pi_i(\theta_i), \quad i = 0, 1. \quad (5)$$

Given any realization X of \mathcal{X} we would like to decide between the two hypotheses H_0, H_1 ; and if our

decision is in favor of H_i , then we would like to provide an estimate $\hat{\theta}_i$ for the corresponding parameters θ_i .

The priors $\pi_i(\theta_i)$ are considered to be generalized functions containing possible point masses. This will allow for the unified analysis of the problem with θ_i taking a continuum or a discrete set of values. Let us now define what we mean by combined detection/estimation scheme.

A. Combined Detection/Estimation Structure

We adopt the class of *randomized* detectors and estimators, and we propose the following *two-step scheme*: In the first step with the help of two randomization probabilities $\delta_0(X), \delta_1(X)$ we decide between H_0, H_1 . Quantity $\delta_i(X)$ denotes the probability by which we decide $d = i$ using a random game. Clearly $\delta_0(X) + \delta_1(X) = 1$. In the second step we provide parameter estimates that we generate with the help of randomized estimators. Specifically we define two conditional pdfs $q_0(\hat{\theta}_0|X)$ and $q_1(\hat{\theta}_1|X)$, that satisfy $\int q_0(\hat{\theta}_0|X) d\hat{\theta}_0 = \int q_1(\hat{\theta}_1|X) d\hat{\theta}_1 = 1$. These two density functions are applied as follows: if in the first step we decide $d = i$, then in the second step we use the pdf $q_i(\hat{\theta}_i|X)$ to generate a random variable $\hat{\theta}_i$ distributed according to $q_i(\hat{\theta}_i|X)$. This variable constitutes our estimate. Randomized estimators are the direct analog of randomized tests used in hypothesis testing and are not uncommon in Bayesian approaches, as one can verify by consulting [8, page 65].

We should note that $q_i(\hat{\theta}_i|X)$ must have the same support as the prior $\pi_i(\theta_i)$ since we expect our estimate $\hat{\theta}_i$ to assume the same values as the true parameter θ_i . This is particularly important if θ_i can take only a finite number of values, in which case $\pi_i(\theta_i)$ and $q_i(\hat{\theta}_i|X)$ will be comprised of point masses. In the latter case, it is easy to see, that we can carry out the analysis using only probabilities instead of pdfs and replace integrals over θ_i and $\hat{\theta}_i$ with sums.

Summarizing: the combined detection/estimation structure is comprised of the two probabilities $\delta_0(X), \delta_1(X)$ (used in the first step to distinguish between the two hypotheses H_0, H_1) and of the two pdfs $q_0(\hat{\theta}_0|X), q_1(\hat{\theta}_1|X)$ (used to provide the necessary parameter estimate in the second step). We denote the complete detection/estimation structure as $\mathcal{D} = \{\delta_0(X), \delta_1(X), q_0(\hat{\theta}_0|X), q_1(\hat{\theta}_1|X)\}$.

Remark 1: One might wonder if the adoption of a two-step procedure covers all possibilities for a randomized detector/estimator. It turns out that we could also use one-step detectors/estimators that *simultaneously* detect and estimate. However, it is straightforward to show that such schemes can be simulated by properly selected two-step procedures; furthermore, the opposite is also true, that is, any two-step detector/estimator can be simulated by a proper one-step procedure. Consequently the two approaches are fully equivalent and, without loss of generality, we may limit ourselves to the two-step

schemes introduced above¹.

In the next subsection our aim is to define a suitable performance measure for \mathcal{D} and a corresponding optimization problem that will lead to the identification of the optimum detection/estimation structure.

B. Combined Optimization Problem

As we mentioned in the Introduction, we are going to combine the Bayesian with the Neyman-Pearson approach. To this end let $C_{ji}(\hat{\theta}_j, \theta_i)$ denote the cost of deciding in favor of hypothesis H_j in the first step and providing the estimate $\hat{\theta}_j$ in the second step, when the true hypothesis is H_i and the true parameter is θ_i .

Let us consider the average cost $\mathcal{C}_i(\mathcal{D})$ given that the true hypothesis is H_i . We can express $\mathcal{C}_i(\mathcal{D})$ in terms of the complete detection/estimation structure as follows

$$\mathcal{C}_i(\mathcal{D}) = \int \left\{ \delta_0(X) \int q_0(\hat{\theta}_0|X) \mathcal{D}_{0i}(\hat{\theta}_0, X) d\hat{\theta}_0 + \delta_1(X) \int q_1(\hat{\theta}_1|X) \mathcal{D}_{1i}(\hat{\theta}_1, X) d\hat{\theta}_1 \right\} dX, \quad (6)$$

where $\mathcal{D}_{ji}(U, X) = \int C_{ji}(U, \theta_i) f_i(X|\theta_i) \pi_i(\theta_i) d\theta_i$. As we can see the four functions $\mathcal{D}_{ji}(U, X)$ depend on the known cost functions $C_{ji}(U, \theta_i)$ and on prior information, consequently they are also known and independent from the detection/estimation structure \mathcal{D} .

We can now define the following optimization problem that we propose as an alternative to the classical problem depicted in (2).

$$\inf_{\mathcal{D}} \mathcal{C}_1(\mathcal{D}), \quad \text{subject to } \mathcal{C}_0(\mathcal{D}) \leq \alpha. \quad (7)$$

Level α constitutes the maximally allowable cost under hypothesis H_0 . As we can see by direct comparison with (2), we follow a Neyman-Pearson like approach, having replaced the (conditional) error probabilities of the classical approach with the conditional Bayesian costs. The problem defined in (7) makes a lot of sense. Indeed if one is interested in parameter estimation under each hypothesis then the primal concern is the induced average estimation cost, which quantifies the quality of the corresponding estimate. It is therefore understandable that both, the detection and the estimation subproblems must contribute towards the optimization of the same figure of merit.

Before continuing with the general solution of our problem, we would like to consider a special case which establishes finite-sample-size optimality for the GLR test. The practical significance of this popular test certainly justifies this special analysis. There is however an additional reason that makes this short parenthesis necessary: we plan to use the GLR test as our prototype, therefore we will observe under

¹Our claim is particularly easy to prove when the parameters θ_i take only a finite number of values.

what conditions we can guarantee its optimality. Then we will apply similar assumptions in the general case, in order to generate GLR-like tests that are compatible with various well known cost functions used in applications. This will produce novel tests that are hopefully more suitable than the classical GLR test, for these problems.

C. Optimality of the GLR Test

Consider the case where θ_i takes a finite set of values. Without loss of generality, we will assume that $\theta_i = 1, 2, \dots, L_i$ and for simplicity, when $\theta_i = l$, we are going to denote the corresponding pdf as $f_{il}(X)$ instead of $f_i(X|\theta_i = l)$. This immediately suggests that the two prior pdfs $\pi_i(\theta_i)$ will be comprised of an equivalent number of point masses. We denote the corresponding prior probabilities with π_{il} . In other words under hypothesis H_i we have $\mathcal{X} \sim f_{il}(X)$ with prior probability π_{il} , where $i = 0, 1$ and $l = 1, \dots, L_i$. Since θ_i assumes a finite number of values, the estimators $q_i(\hat{\theta}_i|X)$ will be comprised of point masses as well. Let $q_{il}(X)$ denote the corresponding probabilities. Our detection/estimation structure can then be identified as the following collection of probabilities

$$\mathcal{D} = \{\delta_0(X), \delta_1(X), q_{01}(X), \dots, q_{0L_0}(X), q_{11}(X), \dots, q_{1L_1}(X)\} \quad (8)$$

with the following properties

$$\delta_i(X) \geq 0; \quad q_{il}(X) \geq 0; \quad \delta_0(X) + \delta_1(X) = \sum_{l=1}^{L_i} q_{il}(X) = 1. \quad (9)$$

As before the probabilities $\delta_0(X), \delta_1(X)$ are used in the first step to decide between the two main hypotheses. Given that the decision in the first step is in favor of H_i , we go to the second step and with the help of the probabilities $q_{il}(X)$, $l = 1, \dots, L_i$, we decide with the help of a randomized test among the possibilities $f_{i1}(X), \dots, f_{iL_i}(X)$.

Consider now the following special case of cost functions

$$C_{10}(\hat{\theta}_1, \theta_0) = C_{01}(\hat{\theta}_0, \theta_1) = 1; \quad C_{11}(\hat{\theta}, \theta) = C_{00}(\hat{\theta}, \theta) = \mathbb{1}_{\{\hat{\theta} \neq \theta\}}, \quad (10)$$

where $\mathbb{1}_A$ denotes the indicator of the set A . In other words the cost is 0 only when both steps make the correct selection and it is equal to 1 otherwise. The corresponding average cost $\mathcal{C}_i(\mathcal{D})$ is then equal to the probability of *detection/estimation-error* under hypothesis H_i . We have the following theorem that solves the problem defined in (7).

Theorem 1. Consider the class \mathcal{J}_α of all detection/estimation strategies that satisfy the constraint

$$P(\text{Detection/estimation-error}|H_0) \leq \alpha, \quad (11)$$

where $\alpha_{\min} < \alpha < 1$, with

$$\alpha_{\min} = 1 - \int \max_{1 \leq l \leq L_0} \{\pi_{0l} f_{0l}(X)\} dX. \quad (12)$$

The test, within the class \mathcal{J}_α , that minimizes the probability $P(\text{Detection/estimation-error}|\mathbf{H}_1)$ is given by:

Step 1: The optimum strategy for deciding between the two main hypotheses \mathbf{H}_0 and \mathbf{H}_1 is

$$\frac{\max_{1 \leq l \leq L_1} \{\pi_{1l} f_{1l}(X)\}}{\max_{1 \leq l \leq L_0} \{\pi_{0l} f_{0l}(X)\}} \underset{\mathbf{H}_0}{\overset{\mathbf{H}_1}{\geq}} \lambda \quad (13)$$

where, whenever the left hand side coincides with the threshold we perform a randomization between the two hypotheses and select \mathbf{H}_1 with probability γ .

Step 2: If in Step 1 we decide in favor of hypothesis \mathbf{H}_j then the optimum estimation strategy is

$$\hat{\theta}_j = \arg \max_{1 \leq l \leq L_j} \{\pi_{jl} f_{jl}(X)\}. \quad (14)$$

If more than one indexes attain the same maximum we perform an arbitrary randomization among them.

The threshold λ and the randomization probability γ of Step 1 must be selected so that the constraint in (11) is satisfied with equality.

Proof: We observe that $P(\text{Detection/estimation-error}|\mathbf{H}_i) = 1 - P(\text{Correct-detection/estimation}|\mathbf{H}_i)$, therefore the constraint is equivalent to $P(\text{Correct-detection/estimation}|\mathbf{H}_0) \geq 1 - \alpha$. If we denote the possibility $\{\mathcal{X} \sim f_{il}(X)\}$ with \mathbf{H}_{il} then we can write

$$P(\text{Correct-detection/estimation}|\mathbf{H}_i) = \sum_{l=1}^{L_i} P(\text{Correct-detection/estimation}|\mathbf{H}_{il}) \pi_{il} \quad (15)$$

with

$$P(\text{Correct-detection/estimation}|\mathbf{H}_{il}) = \int \delta_i(X) q_{il}(X) f_{il}(X) dX. \quad (16)$$

Instead of minimizing the probability of detection/estimation-error we can equivalently maximize the probability of correct-detection/estimation. To solve the constrained optimization problem, let $\lambda > 0$ be a Lagrange multiplier and, as in the classical Neyman-Pearson case, with the help of (15) and (16), define

the corresponding unconstrained version. We then note

$$\begin{aligned} & P(\text{Correct-detection/estimation}|\mathbf{H}_1) + \lambda P(\text{Correct-detection/estimation}|\mathbf{H}_0) \\ &= \int \delta_1(X) \left\{ \sum_{l=1}^{L_1} q_{1l}(X) \pi_{1l} f_{1l}(X) \right\} dX + \lambda \int \delta_0(X) \left\{ \sum_{l=1}^{L_0} q_{0l}(X) \pi_{0l} f_{0l}(X) \right\} dX \end{aligned} \quad (17)$$

$$\leq \int \delta_1(X) \max_{1 \leq l \leq L_1} \{ \pi_{1l} f_{1l}(X) \} dX + \lambda \int \delta_0(X) \max_{1 \leq l \leq L_0} \{ \pi_{0l} f_{0l}(X) \} dX \quad (18)$$

$$= \int \left[\delta_1(X) \max_{1 \leq l \leq L_1} \{ \pi_{1l} f_{1l}(X) \} + \delta_0(X) \lambda \max_{1 \leq l \leq L_0} \{ \pi_{0l} f_{0l}(X) \} \right] dX \quad (19)$$

$$\leq \int \max \left\{ \max_{1 \leq l \leq L_1} \{ \pi_{1l} f_{1l}(X) \}, \lambda \max_{1 \leq l \leq L_0} \{ \pi_{0l} f_{0l}(X) \} \right\} dX. \quad (20)$$

Inequality (18) is valid because the functions $q_{il}(X)$, $l = 1, \dots, L_i$ are nonnegative and complementary (their sum is equal to 1). Inequality (20) is also true because the same properties hold for $\delta_i(X)$, $i = 0, 1$. Note that the final expression constitutes an upper bound on the performance of any detection/estimation rule. Furthermore this upper bound is attainable by a specific detection/estimation strategy. Indeed we note that we have equality in (18) when the estimation probabilities are selected as

$$q_{ik}(X) = \begin{cases} 1 & \text{if } k = \arg \min_{1 \leq l \leq L_i} \{ \pi_{il} f_{il}(X) \} \\ 0 & \text{otherwise,} \end{cases} \quad (21)$$

and we randomize if there are more than one indexes attaining the same maximum. This optimum estimation process is the randomized equivalent of (14). Similarly we have equality in (20) when we select the detection probabilities to be

$$\delta_1(X) = \begin{cases} 1 & \text{if } \max_{1 \leq l \leq L_1} \{ \pi_{1l} f_{1l}(X) \} \geq \lambda \max_{1 \leq l \leq L_0} \{ \pi_{0l} f_{0l}(X) \} \\ \gamma & \text{if } \max_{1 \leq l \leq L_1} \{ \pi_{1l} f_{1l}(X) \} = \lambda \max_{1 \leq l \leq L_0} \{ \pi_{0l} f_{0l}(X) \} \\ 0 & \text{otherwise,} \end{cases} \quad (22)$$

and $\delta_0(X) = 1 - \delta_1(X)$. Clearly this optimum detection procedure is the equivalent of (13).

As far as the false alarm constraint is concerned let us define the following sets

$$\begin{aligned} \mathcal{A}(\lambda) &= \left\{ X : \frac{\max_{1 \leq l \leq L_1} \{ \pi_{1l} f_{1l}(X) \}}{\max_{1 \leq l \leq L_0} \{ \pi_{0l} f_{0l}(X) \}} > \lambda \right\} \\ \mathcal{B}(\lambda) &= \left\{ X : \frac{\max_{1 \leq l \leq L_1} \{ \pi_{1l} f_{1l}(X) \}}{\max_{1 \leq l \leq L_0} \{ \pi_{0l} f_{0l}(X) \}} = \lambda \right\}. \end{aligned} \quad (23)$$

For the test introduced above, we can then write that

$$\begin{aligned}
& P(\text{Detection/estimation-error}|\mathbf{H}_0) \\
&= 1 - \int_{\mathcal{A}(\lambda)} \max_{1 \leq l \leq L_0} \{\pi_{0l} f_{0l}(X)\} dX - \gamma \int_{\mathcal{B}(\lambda)} \max_{1 \leq l \leq L_0} \{\pi_{0l} f_{0l}(X)\} dX \\
&\geq 1 - \int_{\mathcal{A}(\lambda) \cup \mathcal{B}(\lambda)} \max_{1 \leq l \leq L_0} \{\pi_{0l} f_{0l}(X)\} dX \\
&\geq 1 - \int \max_{1 \leq l \leq L_0} \{\pi_{0l} f_{0l}(X)\} dX = \alpha_{\min}.
\end{aligned} \tag{24}$$

The lower bound α_{\min} is clearly attainable in the limit by selecting $\gamma = 1$ and letting $\lambda \rightarrow 0$. Also the detection/estimation-error probability is bounded from above by 1 and we can see that this value can also be attained in the limit by selecting $\gamma = 0$ and letting $\lambda \rightarrow \infty$. Existence of a suitable threshold λ and a randomization probability γ that assure validity of the false alarm constraint with equality, as well as, optimality of the resulting test in the desired sense, can be easily demonstrated following exactly the same steps as in the classical Neyman-Pearson case². This concludes the proof. ■

We realize that in order to apply the test in (13) we need knowledge of the prior probabilities π_{il} . Whenever this information is not available we can consider equiprobable subcases and select $\pi_{il} = 1/L_i$. Under this assumption the optimum test in (13) is reduced to the familiar form of the GLR test,

$$\frac{\max_{1 \leq l \leq L_1} f_{1l}(X)}{\max_{1 \leq l \leq L_0} f_{0l}(X)} \underset{\mathbf{H}_0}{\overset{\mathbf{H}_1}{\geq}} \lambda, \tag{25}$$

after absorbing the two prior probabilities inside the threshold.

Finally, we should mention that if hypothesis \mathbf{H}_0 is simple or, if under hypothesis \mathbf{H}_0 we are not interested in the estimation problem (therefore we can treat it as simple by forming the marginal density) then $P(\text{Detection/estimation-error}|\mathbf{H}_0)$ becomes the usual false alarm probability with corresponding $\alpha_{\min} = 0$. In other words the false alarm probability can take any value in the interval $(0, 1)$ as in the classical Neyman-Pearson problem.

Remark 2: We observe that the optimum test, under each main hypothesis, selects the most appropriate subcase with the help of the MAP selection rule (14). The interesting point is that this selection is performed independently of the other hypothesis and of the corresponding detection strategy. This is clearly a very desirable characteristic since it separates the estimation from the detection problem. In our

²In the proof we simply replace the pdfs $f_i(X)$ with the functions $\max_{1 \leq l \leq L_i} \{\pi_{il} f_{il}(X)\}$. Even though these functions are not densities, the proof goes through without change.

analysis we are going to provide sufficient conditions that can guarantee the same property under the general formulation.

Remark 3: We obtain the GLR test by assuming that the prior probabilities are *uniform*. We will use the same principle in our general formulation to obtain tests that can be used as alternatives to the classical GLR test.

III. OPTIMUM DETECTION/ESTIMATION SCHEME

Let us now continue with the solution of the optimization problem defined in (7). We have the following theorem that provides the desired optimal detection/estimation structure.

Theorem 2. *Consider the class \mathcal{J}_α of detection/estimation structures \mathcal{D} that satisfy $\mathcal{C}_0(\mathcal{D}) \leq \alpha$. The test that minimizes the average cost $\mathcal{C}_1(\mathcal{D})$ within the class \mathcal{J}_α is given by*

$$\inf_U [\mathcal{D}_{01}(U, X) + \lambda \mathcal{D}_{00}(U, X)] \underset{H_0}{\overset{H_1}{\geq}} \inf_U [\mathcal{D}_{11}(U, X) + \lambda \mathcal{D}_{10}(U, X)] \quad (26)$$

with the optimum estimators defined by

$$\hat{\theta}_j = \arg \inf_U [\mathcal{D}_{j1}(U, X) + \lambda \mathcal{D}_{j0}(U, X)], \quad j = 0, 1, \quad (27)$$

and $\lambda > 0$ a threshold properly selected to satisfy the corresponding constraint with equality.

Proof: Let $\lambda > 0$ be a Lagrange multiplier and consider the unconstrained minimization of the combination $\mathcal{C}_1(\mathcal{D}) + \lambda \mathcal{C}_0(\mathcal{D})$. Using (6) we can write

$$\begin{aligned} & \mathcal{C}_1(\mathcal{D}) + \lambda \mathcal{C}_0(\mathcal{D}) \\ &= \int \left\{ \delta_0(X) \int q_0(\hat{\theta}_0|X) [\mathcal{D}_{01}(\hat{\theta}_0, X) + \lambda \mathcal{D}_{00}(\hat{\theta}_0, X)] d\hat{\theta}_0 \right. \\ & \quad \left. + \delta_1(X) \int q_1(\hat{\theta}_1|X) [\mathcal{D}_{11}(\hat{\theta}_1, X) + \lambda \mathcal{D}_{10}(\hat{\theta}_1, X)] d\hat{\theta}_1 \right\} dX \end{aligned} \quad (28)$$

$$\geq \int \left\{ \delta_0(X) \inf_U [\mathcal{D}_{01}(U, X) + \lambda \mathcal{D}_{00}(U, X)] + \delta_1(X) \inf_U [\mathcal{D}_{11}(U, X) + \lambda \mathcal{D}_{10}(U, X)] \right\} dX \quad (29)$$

$$\geq \int \min \left\{ \inf_U [\mathcal{D}_{01}(U, X) + \lambda \mathcal{D}_{00}(U, X)], \inf_U [\mathcal{D}_{11}(U, X) + \lambda \mathcal{D}_{10}(U, X)] \right\} dX. \quad (30)$$

The inequality in (29) is true because

$$\begin{aligned} \int q_i(\hat{\theta}_i|X) [\mathcal{D}_{i1}(\hat{\theta}_i, X) + \lambda \mathcal{D}_{i0}(\hat{\theta}_i, X)] d\hat{\theta}_i &\geq \inf_U [\mathcal{D}_{i1}(U, X) + \lambda \mathcal{D}_{i0}(U, X)] \int q_i(\hat{\theta}_i|X) d\hat{\theta}_i \\ &= \inf_U [\mathcal{D}_{i1}(U, X) + \lambda \mathcal{D}_{i0}(U, X)] \end{aligned} \quad (31)$$

with equality iff $q_i(\hat{\theta}_i|X)$ puts all its probability mass on the choice $\hat{\theta}_i = \arg \inf_U [\mathcal{D}_{i1}(U, X) + \lambda \mathcal{D}_{i0}(U, X)]$, which is thereby optimum. Similarly we have that (30) is true because $\delta_0(X) + \delta_1(X) = 1$, and we have equality iff

$$\delta_1(X) = \begin{cases} 1 & \text{if } \inf_U [\mathcal{D}_{01}(U, X) + \lambda \mathcal{D}_{00}(U, X)] > \inf_U [\mathcal{D}_{11}(U, X) + \lambda \mathcal{D}_{10}(U, X)] \\ \gamma & \text{if } \inf_U [\mathcal{D}_{01}(U, X) + \lambda \mathcal{D}_{00}(U, X)] = \inf_U [\mathcal{D}_{11}(U, X) + \lambda \mathcal{D}_{10}(U, X)] \\ 0 & \text{if } \inf_U [\mathcal{D}_{01}(U, X) + \lambda \mathcal{D}_{00}(U, X)] < \inf_U [\mathcal{D}_{11}(U, X) + \lambda \mathcal{D}_{10}(U, X)], \end{cases} \quad (32)$$

with $0 \leq \gamma \leq 1$ and $\delta_0(X) = 1 - \delta_1(X)$. This is the randomized version of (26). This completes the proof. \blacksquare

Remark 4: For the level α we have $\alpha_{\min} < \alpha < \alpha_{\max}$. It is possible to come up with an expression for α_{\min} . Indeed, from (6) it is easy to see that

$$\mathcal{C}_0(\mathcal{D}) \geq \int \left\{ \delta_0(X) \inf_U \mathcal{D}_{00}(U, X) + \delta_1(X) \inf_U \mathcal{D}_{10}(U, X) \right\} dX \quad (33)$$

$$\geq \int \min \left\{ \inf_U \mathcal{D}_{00}(U, X), \inf_U \mathcal{D}_{10}(U, X) \right\} dX = \alpha_{\min}. \quad (34)$$

This lower bound is in fact attainable by the *optimum scheme* defined with (26), (27), if we let $\lambda \rightarrow 0$. Unfortunately a similar expression for the upper bound α_{\max} was not possible to obtain.

Remark 5: As we can see from (26), (27) the optimal solutions for the detection and estimation subproblems are interrelated. If we are interested in the same characteristic we encountered in the GLR test, where the two estimation problems are independent from each other and from the detection part, then the following special form of the cost functions can assure the validity of this property

$$C_{01}(U, \theta_1) = C_{01}(\theta_1) \text{ and } C_{10}(U, \theta_0) = C_{10}(\theta_0). \quad (35)$$

Indeed we can see that if (35) is true then $\mathcal{D}_{01}(U, X) = \mathcal{D}_{01}(X)$ and $\mathcal{D}_{10}(U, X) = \mathcal{D}_{10}(X)$, which implies that the optimum estimators in (27) simplify to

$$\hat{\theta}_0 = \arg \inf_U [\mathcal{D}_{01}(X) + \lambda \mathcal{D}_{00}(U, X)] = \arg \inf_U \mathcal{D}_{00}(U, X) \quad (36)$$

$$\hat{\theta}_1 = \arg \inf_U [\mathcal{D}_{11}(U, X) + \lambda \mathcal{D}_{10}(X)] = \arg \inf_U \mathcal{D}_{11}(U, X), \quad (37)$$

that is, they coincide with the classical Bayesian estimators which we obtain by treating each estimation problem separately. The optimum detector in (26), under the same assumptions takes the form

$$\mathcal{D}_{01}(X) - \inf_U \mathcal{D}_{11}(U, X) \underset{H_0}{\overset{H_1}{\geq}} \lambda \left[\mathcal{D}_{10}(X) - \inf_U \mathcal{D}_{00}(U, X) \right], \quad (38)$$

which of course relies on the optimum cost values.

Remark 6: Observing (26) and (27) it seems as if the order of the two steps in our two-step procedure has been reversed. This impression however is not exactly correct. We note that the minimum of a function is unique and it is the two minimal values that are used in (26). The actual estimates that realize the two minima, and are depicted in (27), are not necessarily unique and therefore we might require randomization which is performed in the second step. But even if the two estimators are deterministic, it is the first step that will dictate which of the two values will be used as our actual parameter estimate. And this selection is performed *after* the detection step. Therefore, strictly speaking, the order is not reversed.

A. Special Case

We would like now to pay attention to a particular case that is common in applications. Consider under H_1 that $\mathcal{X} \sim f_1(X|\theta)$ where θ a parameter vector with known prior $\pi(\theta)$ and under H_0 we assume that $\mathcal{X} \sim f_0(X)$. In other words the pdf under H_0 is completely known. In fact it is very common to have $f_0(X) = f_1(X|\theta = 0)$. Our goal is to test H_0 against H_1 , and whenever we decide in favor of H_1 to provide an estimate $\hat{\theta}$ for the corresponding parameter vector θ . We should mention that the two application problems discussed in the Introduction, fall under this particular class.

Since parameter estimation is needed only under H_1 , this suggests that a combined detection/estimation structure will be comprised of the following functions $\mathcal{D} = \{\delta_0(X), \delta_1(X), q_1(\hat{\theta}|X)\}$ that satisfy $\delta_j(X) \geq 0$, $j = 0, 1$, $q_1(\hat{\theta}|X) \geq 0$, $\delta_0(X) + \delta_1(X) = \int q_1(\hat{\theta}|X)d\hat{\theta} = 1$. The two probabilities $\delta_0(X), \delta_1(X)$ will be used in the first step to decide between the two main hypotheses, while $q_1(\hat{\theta}|X)$ will be employed in the second step to provide the necessary estimate for θ , every time a decision in favor of H_1 is reached.

Regarding the estimation costs we have the following functions $C_{11}(\hat{\theta}, \theta)$, $C_{10}(\hat{\theta})$, $C_{01}(\theta)$ and C_{00} . As we can see C_{00} is simply a constant, whereas $C_{10}(\cdot)$ and $C_{01}(\cdot)$ are functions of a single quantity. Consider now the following selection $C_{00} = 0$ and $C_{10}(\hat{\theta}) = 1$, then it is easy to verify that $\mathcal{C}_0(\mathcal{D}) = P(d = 1|H_0)$, i.e. the probability of false alarm. For this particular selection we have the following interesting corollary of Theorem 2.

Corollary 1. *Consider the average cost $\mathcal{C}_1(\mathcal{D})$ under H_1 defined using the two cost functions $C_{11}(\hat{\theta}, \theta)$ and $C_{01}(\theta)$. The optimum detection/estimation structure that minimizes $\mathcal{C}_1(\mathcal{D})$ under the constraint that the false alarm probability $P(d = 1|H_0)$ is no larger than $\alpha \in (0, 1)$, is given by*

$$\frac{\mathcal{D}_{01}(X) - \inf_U \mathcal{D}_{11}(U, X)}{f_0(X)} \underset{H_0}{\overset{H_1}{\leq}} \lambda, \quad (39)$$

for the optimum detector and

$$\hat{\theta}_1 = \arg \inf_U \mathcal{D}_{11}(U, X) \quad (40)$$

for the corresponding optimum estimator. The two functions $\mathcal{D}_{11}(U, X)$, $\mathcal{D}_{01}(X)$ are defined as follows

$$\mathcal{D}_{11}(U, X) = \int C_{11}(U, \theta) f_1(X|\theta) \pi(\theta) d\theta; \quad \mathcal{D}_{01}(X) = \int C_{01}(\theta) f_1(X|\theta) \pi(\theta) d\theta. \quad (41)$$

B. Discussion

In finite-sample-size optimum detection and estimation the need for the prior pdfs constitutes a very severe weakness. As we mentioned earlier, if this information is not available the corresponding optimization problems must be treated in some min-max context. Unfortunately min-max formulations tend to be very difficult to solve even asymptotically, and no systematic solution exists for the problems of detection and estimation. It is of course clear that the same limitation applies in the case of the more general combined detection/estimation problem.

A simple (ad-hoc) method to bypass the need for resorting to min-max approaches, is to apply the same idea used to demonstrate optimality for the GLR test, namely assume that the priors are *uniform*. Of course this selection is arbitrary and does not guarantee optimality of the corresponding scheme under any possible min-max sense. On the other hand, it is the only logical choice that reflects our complete lack of knowledge about the priors. The corresponding tests, examples of which will be seen in the next section, it is expected to have the same weakness as the GLR test, with one major difference: they will be tailored to the specific cost function adopted in the estimation subproblem.

IV. EXAMPLES

In this section we present a number of interesting examples by selecting various well known forms of cost functions. We basically concentrate on the popular costs encountered in the classical Bayesian estimation theory. We start with the MAP estimate which demonstrates optimality of the GLR test in the continuous case.

A. MAP Detection/Estimation

Consider the following combination of cost functions

$$C_{01}(U, \theta) = C_{10}(U, \theta) = 1; \quad C_{00}(U, \theta) = C_{11}(U, \theta) = \begin{cases} 0 & \|U - \theta\| \leq \Delta \ll 1 \\ 1 & \text{otherwise.} \end{cases} \quad (42)$$

We recall from the classical Bayesian estimation theory (see [9, Page 145]) that, as $\Delta \rightarrow 0$ and assuming sufficient smoothness of the pdf functions, the specific selection of costs leads to the MAP parameter estimation under each main hypothesis. Indeed we observe³

$$\mathcal{D}_{jj}(U, X) \approx \int f_j(X|\theta)\pi_j(\theta)d\theta - f_j(X|U)\pi_j(U)V_j(\Delta) \quad (43)$$

where $V_j(\Delta)$ is the volume of a hypersphere of radius Δ (which can be different for each hypothesis if the two parameter vectors are not of the same length). Substituting in (38) yields

$$\frac{\sup_U f_1(X|U)\pi_1(U)}{\sup_U f_0(U|X)\pi_0(U)} \underset{H_0}{\overset{H_1}{\cong}} \lambda \frac{V_0(\Delta)}{V_1(\Delta)} = \lambda', \quad (44)$$

and the optimum estimator under each hypothesis is the MAP estimator

$$\hat{\theta}_j = \arg \sup_U f_j(X|U)\pi_j(U). \quad (45)$$

Similarly for the special case of Corollary 1 if we define

$$C_{11}(U, \theta) = \begin{cases} 0 & \|U - \theta\| \leq \Delta \ll 1 \\ 1 & \text{otherwise,} \end{cases} \quad (46)$$

and $C_{01}(\theta) = 1$, then $\mathcal{D}_{11}(U, X) \approx \int f_1(X|\theta)\pi(\theta)d\theta - f_1(X|U)\pi(U)V_1(\Delta)$ and the optimum test in (39) takes the form

$$\frac{\sup_U f_1(X|U)\pi(U)}{f_0(X)} \underset{H_0}{\overset{H_1}{\cong}} \frac{\lambda}{V_1(\Delta)} = \lambda', \quad (47)$$

with the optimum estimator being $\hat{\theta} = \arg \sup_U f(X|U)\pi(U)$. In both tests (44) and (47), the threshold λ' and the corresponding randomization probability γ are selected to satisfy the false alarm constraint with equality. If the prior probabilities $\pi_i(\theta_i), \pi(\theta)$ are unknown and are replaced with the uniform over some prior sets Θ_i we obtain the classical form of the GLR test depicted in (3).

B. MMSE Detection/Estimation

Let us now develop the first test that can be used as an alternative to the GLR test. Consider the following costs

$$C_{01}(U, \theta_1) = C_{01}(\theta_1); \quad C_{10}(U, \theta_0) = C_{10}(\theta_0); \quad C_{00}(U, \theta) = C_{11}(U, \theta) = \|U - \theta\|^2, \quad (48)$$

³The approximate equality becomes exact as $\Delta \rightarrow 0$.

where $C_{01}(\theta_1)$, $C_{10}(\theta_1)$ are functions to be specified in the sequel. Due to the specific form of the costs, the two estimators are independent from each other and also independent from the detection part. Under each main hypothesis the optimum estimator is obtained by minimizing the corresponding mean square error. Consequently the optimum estimator is the conditional mean of the parameter vector given the data vector X (see [9, Page 143]). Specifically we have

$$\hat{\theta}_j = E[\theta_j|X, H_j] = \frac{\int \theta_j f_j(X|\theta_j) \pi_j(\theta_j) d\theta_j}{\int f_j(X|\theta_j) \pi_j(\theta_j) d\theta_j}. \quad (49)$$

The corresponding optimum test after substituting in (38) takes the form

$$\mathcal{A}_1(X) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \mathcal{A}_0(X) \quad (50)$$

where

$$\begin{aligned} \mathcal{A}_0(X) &= \|\hat{\theta}_0\|^2 f_0(X) + \int [C_{10}(\theta_0) - \|\theta_0\|^2] f_0(X|\theta_0) \pi_0(\theta_0) d\theta_0 \\ \mathcal{A}_1(X) &= \|\hat{\theta}_1\|^2 f_1(X) + \int [C_{01}(\theta_1) - \|\theta_1\|^2] f_1(X|\theta_1) \pi_1(\theta_1) d\theta_1 \\ f_j(X) &= \int f_j(X|\theta_j) \pi_j(\theta_j) d\theta_j. \end{aligned} \quad (51)$$

Selecting $C_{01}(\theta_1) = \|\theta_1\|^2$ and $C_{10}(\theta_0) = \|\theta_0\|^2$ simplifies the test considerably yielding

$$\frac{\|\hat{\theta}_1\|^2 f_1(X)}{\|\hat{\theta}_0\|^2 f_0(X)} = \frac{\|\hat{\theta}_1\|^2 \int f_1(X|\theta_1) \pi_1(\theta_1) d\theta_1}{\|\hat{\theta}_0\|^2 \int f_0(X|\theta_0) \pi_1(\theta_0) d\theta_0} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda. \quad (52)$$

We recognize in the second ratio the likelihood that is used to decide optimally between the two main hypotheses. By including the first ratio of the two norm square estimates, the test performs simultaneously optimum detection and estimation.

For the special case of Corollary 1 it is easy to verify that the corresponding test takes the form

$$\frac{\|\hat{\theta}_1\|^2 f_1(X) + \int [C_{01}(\theta) - \|\theta\|^2] f_1(X|\theta) \pi(\theta) d\theta}{f_0(X)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda, \quad (53)$$

which, if we select $C_{01}(\theta) = \|\theta\|^2$, simplifies to

$$\|\hat{\theta}\|^2 \frac{f_1(X)}{f_0(X)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda, \quad (54)$$

where $\hat{\theta} = E[\theta|X, H_1] = \int \theta f(X|\theta) \pi(\theta) d\theta / \int f(X|\theta) \pi(\theta) d\theta$ and $f_1(X) = \int f(X|\theta) \pi(\theta) d\theta$.

In both tests in (50) and (53), if the priors are not known and are replaced by uniforms, we obtain tests that are the equivalent of the GLR test for the MMSE criterion.

C. Median Detection/Estimation

As our final example we present the case of the median estimation where $\theta_i, \hat{\theta}_i, \theta, U$ are scalars and we select the cost functions as follows

$$C_{01}(U, \theta) = C_{01}(\theta); \quad C_{10}(U, \theta) = C_{10}(\theta); \quad C_{00}(U, \theta) = C_{11}(U, \theta) = |U - \theta|. \quad (55)$$

The estimators are again independent from each other and from detection part. Under each hypothesis we perform optimum Bayes estimation and for this specific cost function we know that the optimum estimator is the conditional median [9, Page 143]

$$\hat{\theta}_j = \arg \left\{ y : P(\theta_j \leq y | X, H_j) = \frac{\int_{-\infty}^y f_j(X|\theta_j)\pi_j(\theta_j) d\theta_j}{\int f_j(X|\theta_j)\pi_j(\theta_j) d\theta_j} = \frac{1}{2} \right\}. \quad (56)$$

The optimum test, as before, becomes

$$\mathcal{A}_1(X) \underset{H_0}{\overset{H_1}{\geq}} \lambda \mathcal{A}_0(X) \quad (57)$$

where

$$\begin{aligned} \mathcal{A}_0(X) &= \int \left[C_{10}(\theta_0) + \theta_0 \text{sgn}(\hat{\theta}_0 - \theta_0) \right] f_0(\theta_0|X) \pi_0(\theta_0) d\theta_0 \\ \mathcal{A}_1(X) &= \int \left[C_{01}(\theta_1) + \theta_1 \text{sgn}(\hat{\theta}_1 - \theta_1) \right] f_1(\theta_1|X) \pi_1(\theta_1) d\theta_1. \end{aligned} \quad (58)$$

If additionally we select $C_{01}(\theta_1) = |\theta_1|$ and $C_{10}(\theta_0) = |\theta_0|$ then the optimum test takes the more convenient form

$$\frac{\int_0^{\hat{\theta}_1} \theta_1 f_1(X|\theta_1) \pi_1(\theta_1) d\theta_1}{\int_0^{\hat{\theta}_0} \theta_0 f_0(X|\theta_0) \pi_0(\theta_0) d\theta_0} \underset{H_0}{\overset{H_1}{\geq}} \lambda. \quad (59)$$

For the special case of Corollary 1 and for $C_{01}(\theta) = |\theta|$, the corresponding optimum test reduces to

$$\frac{\int_0^{\hat{\theta}} \theta f(X|\theta) \pi(\theta) d\theta}{f_0(X)} \underset{H_0}{\overset{H_1}{\geq}} \lambda, \quad (60)$$

while the optimum estimator is $\hat{\theta} = \arg\{y : P(\theta \leq y | X, H_1) = 0.5\}$. Finally when the priors are selected to be uniform, we obtain a test that is the alternative to the GLR test but tuned to the specific Bayesian criterion we employ in the estimation part.

V. APPLICATION TO RETROSPECTIVE CHANGEPOINT DETECTION

Perhaps the most appropriate application where one would readily need to replace the GLR test with an alternative scheme, is the problem of target detection and localization. Clearly for this problem the most suitable cost function is the mean square error between the location estimate and the true position. This choice will inevitably lead to the use of tests that are similar to (53), proposing a completely novel approach for this intriguing problem. Unfortunately the corresponding derivations are lengthy and thus impossible to detail here. In the limiting space we have to our disposal it is feasible to treat, with our preceding methodology, the second application we mentioned in the Introduction, namely the retrospective changepoint detection problem. We would like to mention that even though in this problem the estimation costs are MAP-like, suggesting use of the GLR test, as we will see, there is sufficient simplicity and originality in our results that make our analysis interesting and worth including in this article.

In its *simplest* form, retrospective changepoint detection is about an observation vector $\mathcal{X} \in \mathbb{R}^N$ and two pdfs $f_\infty(X)$ and $f_0(X)$ which are completely known. If $\mathcal{X} = [\chi_1, \dots, \chi_N]^t$ then we assume that there is an unknown point τ such that the samples $\{\chi_1, \dots, \chi_\tau\}$ follow the nominal measure $f_\infty(X)$ while the $\{\chi_{\tau+1}, \dots, \chi_N\}$ switch to the alternative $f_0(X)$. Consequently, the changepoint τ is the *last* point where the samples follow the nominal regime⁴.

We are interested in deciding whether the change took place within or before the given collection of samples, that is $\tau < N$, or the change will take place at some future point (possibly at infinity), that is $\tau \geq N$. In the former case we would also like to obtain an estimate $\hat{\tau}$ of the changepoint τ . The combined detection/estimation version of the retrospective changepoint detection problem, as we mentioned in the Introduction, is suitable for formulating segmentation problems.

Let us first define the joint pdf $f_\tau(X)$ of the samples \mathcal{X} given τ . We distinguish three sets of values for τ , namely $\tau \leq 0$, $\tau \in \{1, \dots, N-1\}$ and $\tau \geq N$. The first corresponds to a change occurring before taking any samples, the second to a change within the available sample set and the third to the change occurring after we acquired the samples. The most common model for the induced joint pdf is [10], [11]

$$f_\tau(X) = \begin{cases} f_0(X) & \text{for } \tau \leq 0 \\ f_\infty(X_1^\tau) f_0(X_{\tau+1}^N | X_1^\tau) & \text{for } 0 < \tau \leq N-1 \\ f_\infty(X) & \text{for } N \leq \tau, \end{cases} \quad (61)$$

⁴Notation seems to be somewhat awkward compared to the usual one used in hypothesis testing. We simply follow the standard practice of sequential changepoint detection theory.

where $X = [x_1, \dots, x_N]^t$ and for $a \leq b$ we define $X_a^b = [x_a, \dots, x_b]^t$. We can see that if the change takes place before the samples are acquired, all samples are under the alternative regime. If the change takes place within the available set, then the initial portion of the samples follows the pdf of the nominal regime while the final portion the *conditional* pdf of the alternative regime. Finally if the change does not occur before or inside the available data set, all samples are under the nominal regime.

Regarding the changepoint τ there are different models. Detailed discussion of the various possibilities can be found in [10], [11]. Here we limit ourselves to Shiryaev's popular Bayesian model. Specifically we assume that τ is a random variable with a prior $\{\varpi_n\}$ defined as $\varpi_0 = P(\tau \leq 0)$, $\varpi_n = P(\tau = n)$ for $0 < n \leq N - 1$, $\varpi_N = P(\tau \geq N)$ and such that $\sum_{n=0}^N \varpi_n = 1$.

As we mentioned, the goal is to test $\{\tau \leq N - 1\}$ against $\{\tau \geq N\}$, and in the former case provide and estimate $\hat{\tau}$ for τ . Formulating the problem according to our previous theory, we have that under H_0 the samples follow the nominal pdf $f_\infty(X)$ while under H_1 we have N different possibilities with corresponding pdf $f_\tau(X)$ and prior $\pi_\tau = \varpi_\tau / (\sum_{k=0}^{N-1} \varpi_k) = \varpi_\tau / (1 - \varpi_N)$, where $0 \leq \tau < N$.

Let us now consider the combined detection/estimation problem in the sense of Corollary 1, namely minimize the average cost under H_1 subject to a false alarm probability constraint under H_0 . We propose the following cost functions $C_{11}(\hat{\tau}, \tau) = \mathbb{1}_{\{\hat{\tau} \neq \tau\}}$, where $\mathbb{1}_A$ denotes the indicator function of the set A , and $C_{01}(\tau) = 1$. In other words we penalize with 1 the incorrect detection of H_1 but also the correct detection of H_1 followed by an incorrect estimation of τ . The average cost is simply the probability of detection/estimation-error introduced in Subsection II.C. Applying the results of Corollary 1 and using the Bayes rule and (61), the optimum detection/estimation structure is given by

$$\max_{0 \leq n < N} \left\{ \pi_n \frac{f_n(X)}{f_\infty(X)} \right\} = \max_{0 \leq n < N} \left\{ \pi_n \frac{f_0(X_{n+1}^N | X_1^n)}{f_\infty(X_{n+1}^N | X_1^n)} \right\} \begin{matrix} H_1 \\ \geq \\ \lambda \\ \leq \\ H_0 \end{matrix} \quad (62)$$

for the optimum detector and

$$\hat{\tau} = \arg \max_{0 \leq n < N} \left\{ \pi_n \frac{f_0(X_{n+1}^N | X_1^n)}{f_\infty(X_{n+1}^N | X_1^n)} \right\} \quad (63)$$

for the corresponding optimum estimator. If the priors π_n , $n = 0, \dots, N - 1$, are unknown and we select them to be equal, then we obtain the GLR test version of the problem

$$\mathcal{S}_N = \max_{0 \leq n < N} \left\{ \frac{f_0(X_{n+1}^N | X_1^n)}{f_\infty(X_{n+1}^N | X_1^n)} \right\} \begin{matrix} H_1 \\ \geq \\ \lambda \\ \leq \\ H_0 \end{matrix} \quad (64)$$

where \mathcal{S}_N is known as the *CUSUM statistic* for point N .

The previous result was of course expected since we followed the same formulation as the one used in Subsection II.C to prove optimality of the GLR test. Interestingly, our theory allows for the development of alternative detection/estimation structures in a simple and straightforward manner. For example one might argue that the cost $C_{11}(\hat{\tau}, \tau) = \mathbb{1}_{\{\hat{\tau} \neq \tau\}}$ is overly stringent and propose as alternative the function $C_{11}(\hat{\tau}, \tau) = \mathbb{1}_{\{|\hat{\tau} - \tau| > m\}}$ where $0 \leq m \ll N$ is a nonnegative integer. In other words we tolerate errors in the estimate of τ that do not exceed m points. If $m = 0$ the problem is reduced to the case already discussed. Clearly most practical segmentation problems would allow $m > 0$.

Again we adopt the setup proposed in Corollary 1. It is then easy to verify that we obtain the following optimum structure

$$\max_{m \leq n < N-m} \left\{ \sum_{k=-m}^m \pi_{n+k} \frac{f_{n+k}(X)}{f_{\infty}(X)} \right\} = \max_{m \leq n < N-m} \left\{ \sum_{k=-m}^m \pi_{n+k} \frac{f_0(X_{n+k+1}^N | X_1^{n+k})}{f_{\infty}(X_{n+k+1}^N | X_1^{n+k})} \right\} \begin{matrix} \text{H}_1 \\ \geq \\ \text{H}_0 \end{matrix} \lambda \quad (65)$$

for the detector and

$$\hat{\tau} = \arg \max_{m \leq n < N-m} \left\{ \sum_{k=-m}^m \pi_{n+k} \frac{f_0(X_{n+k+1}^N | X_1^{n+k})}{f_{\infty}(X_{n+k+1}^N | X_1^{n+k})} \right\} \quad (66)$$

for the estimator. Finally assuming uniform priors for the case where the probabilities $\{\pi_0, \dots, \pi_{N-1}\}$ are unknown, leads to the test

$$\bar{S}_N = \max_{m \leq n < N-m} \left\{ \sum_{k=-m}^m \frac{f_0(X_{n+k+1}^N | X_1^{n+k})}{f_{\infty}(X_{n+k+1}^N | X_1^{n+k})} \right\} \begin{matrix} \text{H}_1 \\ \geq \\ \text{H}_0 \end{matrix} \lambda, \quad (67)$$

which is completely novel and replaces the GLR test in (64), with \bar{S}_N being clearly different than the CUSUM statistic.

VI. CONCLUSION

By introducing a joint detection/estimation formulation that properly combines the Neyman-Pearson methodology (for detection) and the Bayesian methodology (for estimation), we derived optimum schemes for problems that require simultaneous detection and estimation. Important side-product of our analysis is the demonstration that the well known GLR test is finite-sample-size optimum under this joint-problem sense. Furthermore we were able to provide completely novel GLR-type tests, that were derived by replacing the MAP estimation cost function with other well known choices as the mean square or mean absolute estimation error. Finally, we used our proposed methodology to analyze the problem of retrospective changepoint detection. This led to the development of a novel detection/estimation structure that can replace the CUSUM approach which is obtained when we apply the GLR test.

ACKNOWLEDGMENT

The author would like to thank his good friend, Prof. Igor Nikiforov from the Université de Technologie de Troyes (UTT), France, for enlightening discussions.

REFERENCES

- [1] I. Bekkerman and J. Tabrikian, "Target detection and localization using MIMO radars and sonars," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3873-3883, 2006.
- [2] N.H. Lehmann, A.M. Haimovich, R.S. Blum, and L. Cimini, "High resolution capabilities of MIMO radar," in Proc. 38th Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, CA, November 2006.
- [3] A. Vexler and C. Wu, "An optimal retrospective change point detection policy," *Scandinavian J. Stat.*, vol. 36, no. 3, pp. 542-558, 2009.
- [4] I. Berkesa, E. Gombay and L. Horváth, "Testing for changes in the covariance structure of linear processes," *J. Stat. Plann. Inference*, vol. 139, pp. 2044-2063, 2009.
- [5] F.C. Robey, et al. "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Systems*, vol.28, no.1, pp.208-216, Jan. 1992.
- [6] S. Bose and A.O. Steinhardt, "A maximal invariant framework for adaptive detection with structured and unstructured covariance matrices," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2164-2175, Sept. 1995.
- [7] S. Bose and A.O. Steinhardt, "Optimum array detector for a weak signal in unknown noise," *IEEE Trans. Aerosp. Electron. Systems*, vol. 32, no. 3, pp.911-922, July 1996.
- [8] C.P. Robert, *The Bayesian Choice*, 2nd edition, Springer, New York, 2007.
- [9] H.V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd edition, Springer, New York, 1994.
- [10] G.V. Moustakides, "Sequential change detection revisited," *Ann. Stat.*, vol. 36, no. 2, pp. 787-807, April 2008.
- [11] G.V. Moustakides, "Change-time models and performance measures for sequential change detection," in Proc. 2nd Int. Workshop Sequential Methodologies, IWSM 2009, Troyes, France, 2009.