

PARAMETRIZED DESIGN OF THE GENERALIZED SEQUENTIAL PROBABILITY RATIO TEST

Naeem Akl, Ahmed Tewfik

Department of Electrical and Computer Engineering
UT Austin

Abstract—A generalized sequential probability ratio test (GSPRT) is a classical algorithm for binary sequential hypothesis testing. Though it is well-studied in the literature, there has been no optimal design of this test due to the difficulty of choosing its thresholds. In this paper we formulate the binary sequential hypothesis testing as an optimization problem. The latter is non-convex, and finding a global minimizer of the objective is combinatorially complex in the number of stages of the sequential test. On the other hand, greedily minimizing the objective has linear complexity but achieves sub-optimal results. We propose a generalization of the greedy approach that allows the designer to trade off complexity for closeness of the thresholds to their optimal values. Simulation results show that the proposed method gives arbitrarily close solution to optimal by increasing the window span of future sample distributions that are utilized to set a current test threshold. The window span is a hyper-parameter that is optimized for the target application.

Index — GSPRT, Sequential Test, Optimization, Average Sample Number, False alarm and Miss Probabilities

I. INTRODUCTION

Sequential hypothesis testing is a statistical method for identifying the true hypothesis within a set of other candidates. Associated to each hypothesis is a random sequence of observations with statistical properties that uniquely mark the hypothesis. A detector evaluates these properties from the observed information sequence and then declares the associated hypothesis to be true. The detector collects more observations in a sequential manner in order to increase the confidence about the true hypothesis and reduce the chance of error in the final decision as desired. As opposed to sequential sampling, batch-mode detection is based on processing a fixed sample size. Sequential detection has the advantage that it reduces the number of collected samples whenever enough confidence is built about the true hypothesis. In addition, sequential hypothesis testing allows to separately optimize for the different types of errors associated with the tested hypotheses.

Since sequential sampling is important for online decision-making, and since minimizing the number of observed information samples is desirable for real-time processing, sequential detection has important applications in economics, engineering and medicine [1]. Unfortunately, optimal sequential testing has only been resolved for independent stationary observations, while in more general settings this is often hindered by the design complexity of the detector. It then becomes critical to push performance more towards optimal in the space where complexity does not skyrocket and as imposed by the particular context in which the detector will function.

Sequential testing is studied in the literature in varied settings. Wald's sequential probability ratio test (SPRT) is the optimal test for independent and identically distributed

observations [2] in binary hypothesis testing. The test thresholds are computed in terms of type I and type II errors and are time-invariant. In [3], asymptotic optimality of the SPRT is proved for the multi-hypothesis version. Optimality is shown for the generalized SPRT (GSPRT) of time-varying thresholds in [4] for the case where the observations are independent but non-identically distributed. This is a generalization of Wald's SPRT. Asymptotic optimality is again established for the GSPRT in [5] in terms of the expected sample size. The GSPRT is shown to be the optimal sequential test for dependent observations in [6]. In none of these works a method to compute the GSPRT thresholds is constructively described. In [7] these thresholds are computed (sub-optimally) using the distributions of the sufficient statistic at each stage of the sequential test disjointly. In a different context where not all the observed samples belong to the same hypothesis, a dynamic programming approach is presented in [8] to compute the thresholds of the generalized Shiriyayev sequential probability ratio test. The test detects a hypothesis change in minimum time.

In this paper we present a general design technique for the GSPRT thresholds for independent non-identically distributed observations. The proposed algorithm allows to dynamically adjust the complexity-accuracy tradeoff of the detector as suited for the target application. A formal statement of the problem including a definition of optimality and the design considerations is presented in Section II. In Section III we highlight the challenges that impede the design process. In Section IV we present an algorithm to compute the GSPRT thresholds and show how greedy and optimal designs are corner cases of the generalized algorithm. Simulation results and insights for three GSPRTs and an SPRT are presented in Section V. Section VI concludes the paper.

II. PROBLEM STATEMENT

Consider a binary hypothesis testing context in which one of two hypotheses H_0 and H_1 is true and should be detected. The detection is based on observing a vector of samples Y_1, \dots, Y_N characterized by probability distribution functions $f_1''(Y_1|H_k), \dots, f_N''(Y_N|H_k)$ that are conditioned on the unknown true hypothesis $H_k, k \in \{0, 1\}$. The samples are presented sequentially to a detector rather than in batch mode. The detector accumulates a sufficient statistic over the collected samples until a decision can be made about the true hypothesis within some error probability bound.

Since observation Y_n could be sampled from one of two possible distributions, the detector computes the corresponding log-likelihood term l_n given by

$$l_n = \log \left(\frac{f_n''(Y_n|H_1)}{f_n''(Y_n|H_0)} \right) \quad (1)$$

Note that l_n is then a function of Y_n , and thus log-likelihood terms $\{l_n\}$ can be described by another set of distributions $f'_1(l_1|H_k), \dots, f'_N(l_N|H_k)$ characteristic of the true hypothesis $H_k, k \in \{0, 1\}$. Observations Y_1, \dots, Y_N are assumed to be independently distributed. Therefore, upon collecting the n^{th} sample Y_n , the cumulative log-likelihood computed by the detector can be expressed as

$$L_n = \sum_{i=1}^n l_i \quad (2)$$

and is distributed according to $f_n(L_n|H_k)$ under H_k . This metric is compared to thresholds a_n and b_n at the n^{th} stage of the sequential step, where the stage index is defined by the number of collected samples so far. If $L_n < a_n$, the detector declares H_0 to be true. If $L_n > b_n$, hypothesis H_1 is declared true. In both cases, the detector no longer collects additional information samples. However, if $a_n < L_n < b_n$, the detector does not decide on the true hypothesis at stage n of the test but collects observation Y_{n+1} . We always refer to the lower and higher thresholds as a_n and b_n respectively. If $g_n(L_n|H_k)$ denotes the distribution of L_n under hypothesis H_k conditioned on $a_n < L_n < b_n$, then it should be clear that

$$f_{n+1}(L_{n+1}|H_k) = g_n(L_{n+1}|H_k) * f'(L_{n+1}|H_k) \quad (3)$$

where $*$ is the convolution operator.

The detector has no control over the distributions of the observations since this is imposed by nature. Thus the binary sequential detector becomes uniquely defined by the sets of thresholds $\{a_n\}$ and $\{b_n\}$, where $1 \leq n \leq N$. There are three main considerations upon defining these thresholds. First, thresholds $\{a_n\}$ should be selected so that the probability that they are traversed by L_n under hypothesis H_1 is minimized. The latter is referred to as *probability of miss* $P_{\mathcal{M}}$. Second, thresholds $\{b_n\}$ on the other hand should be chosen to minimize the *probability of false alarm* $P_{\mathcal{F}}$, which refers to the case where L_n surpasses threshold b_n at some stage n of the sequential test. Since the collected observations and consequently the computed log-likelihood terms are random, the stage of termination of the sequential test N is random. The third consideration is to minimize the *average sample number* ASN, which is the expected value of N under hypothesis $H_k, k \in \{0, 1\}$.

It should be noted that the three considerations are competing. For instance, both $P_{\mathcal{M}}$ and $P_{\mathcal{F}}$ are zero if the thresholds are selected to be infinite. However, in this case the ASN is also infinite. Thus a detector is optimal if its thresholds are chosen so that the ASN is minimized for a given specification of the error probabilities $P_{\mathcal{M}}$ and $P_{\mathcal{F}}$. For identically distributed observations and constant thresholds a_n and b_n , the sequential probability ratio test or SPRT is optimal, in which case $a_n = \log\left(\frac{P_{\mathcal{M}}}{1-P_{\mathcal{F}}}\right)$ and $b_n = \log\left(\frac{1-P_{\mathcal{M}}}{P_{\mathcal{F}}}\right)$. We consider the problem of designing the generalized version of the SPRT where thresholds a_n and b_n may be time-varying and the observations do not necessarily have the same distributions though they are still independent.

III. CHARACTERISTICS OF THE GSPRT DESIGN

We emphasize three characteristics that hinder the design of the GSPRT. Without loss of generality, assume that the

true hypothesis is H_1 . This is unknown to the detector before collecting data. Instead, the detector is fed with the prior probabilities of hypotheses H_0 and H_1 , and these probabilities are equal. Again, the error specifications $P_{\mathcal{M}}$ and $P_{\mathcal{F}}$ are equal. This way the design of the detector thresholds becomes symmetric upon computing these thresholds using distributions $\{f_n(L_n|H_k)\}$ under the two hypotheses H_0 and H_1 . Thus we assume that $a_n = -b_n$, but this can be easily extended to a different setting.

Under H_1 and at stage n of the GSPRT, either a miss, a detection of H_1 or no decision can occur, which happen with probabilities p_n , q_n and $\bar{p}_n = 1 - p_n - q_n$ respectively. Note that these are conditional probabilities, i.e. given that the GSPRT is at the n^{th} stage. Moreover, with the constraint $a_n = -b_n$, p_n and q_n are inter-dependent. While the design of the GSPRT is defined by the selection of the test thresholds, an equivalent set of design parameters is the sequences $\{p_n\}$ and $\{q_n\}$ indexed over n . The ASN under H_1 can be expressed as follows:

$$\text{ASN} = (q_1 + p_1) * 1 + (1 - q_1 - p_1) * (1 + [(q_2 + p_2) * 1 + (1 - q_2 - p_2) * (1 + \dots)]) \quad (4)$$

Note that the ASN is linear in $(q_1 + p_1)$. Therefore, $m_{11} = \frac{\partial^2 \text{ASN}}{\partial(q_1 + p_1)^2} = 0$. Similarly, $m_{22} = \frac{\partial^2 \text{ASN}}{\partial(q_2 + p_2)^2} = 0$. Using the properties of the partial derivatives, we also have $m_{12} = \frac{\partial^2 \text{ASN}}{\partial(q_1 + p_1)\partial(q_2 + p_2)} = \frac{\partial^2 \text{ASN}}{\partial(q_2 + p_2)\partial(q_1 + p_1)} = m_{21}$. Thus,

$$\det \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = -m_{12}^2 \leq 0 \quad (5)$$

However, this determinant is a leading principal minor of the Hessian matrix of the ASN over set $\{p_n + q_n\}$. By Sylvester's criterion the Hessian matrix cannot be positive-definite (nor positive semi-definite), and thus the problem of optimizing objective (4) is non-convex. This sets a challenge in designing the GSPRT.

Another challenge is the computation of the distributions and joint distributions of the sufficient statistic L_n . For instance, the probability that the GSPRT proceeds up to step n under H_0 and outputs H_1 as the true hypothesis at step n is given by $\text{Prob}(a_1 < L_1 < b_1, \dots, a_{n-1} < L_{n-1} < b_{n-1}, L_n > b_n | H_0)$. Finding the distribution of each $\{L_n\}$ as well as their joint distribution is convoluted for general distributions of $\{Y_n\}$. The dependence of the thresholds on such distributions renders finding their optimal values in turn intricate.

The last example illustrates a third important challenge in designing the GSPRT detector. In particular, the values of a_1 and b_1 at step 1 of the test impact the probability of a false alarm at stage $n > 1$. Thus although the test proceeds in a sequential manner, an *optimal design in general* cannot be sequential but requires an overview of the distributions of $\{L_n\}$ over all the test stages.

IV. COMPUTATION OF THE GSPRT THRESHOLDS

In this section we first derive a different form for the ASN than (4). Using the new form we identify two approaches to solve for the thresholds: greedy and optimal. We illustrate the pros and cons of each. Then we show how these two approaches fit within a broader space of solutions for the GSPRT

thresholds. Then we can select an intermediate approach from this space that achieves high performance of the test at low design complexity.

A. Modified Expression for the ASN

Let $t_m = E[N - m | N > m, H_1]$ be the expected number of extra samples to be collected under H_1 before the sequential test terminates given that m samples have already been observed, where $m \geq 0$ and $E[\cdot]$ is the expectation operator. With change of variables, the ASN expression in (4) can be re-defined recursively as

$$\begin{aligned} t_0 &= 1 + \bar{p}_1 * t_1 \\ t_1 &= 1 + \bar{p}_2 * t_2 \\ &\dots \\ t_{M-2} &= 1 + \bar{p}_{M-1} * t_{M-1} \\ t_{M-1} &= 1 \end{aligned} \quad (6)$$

where $M \geq N$ is an arbitrarily large number. In (6), multiply both sides of the first equation by 1, the second equation by \bar{p}_1 , the third equation by $p_1 * p_2, \dots$, and the penultimate equation by $p_1 * p_2 * \dots * p_{M-2}$. Starting from the last equation and substituting each equation in the previous one we obtain

$$\text{ASN} = t_0 = 1 + (\bar{p}_1) + (\bar{p}_1 * \bar{p}_2) + \dots + (\bar{p}_1 * \bar{p}_2 * \dots * \bar{p}_{M-1}) \quad (7)$$

B. Greedy versus Optimal Solutions

Assume only for now that terms $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_{M-1}$ are independent and have no constraints other than being probability measures. Given this assumption, we claim that for $M - 1$ fixed values assigned to elements $\{\bar{p}_m\}_{m=1}^{M-1}$, objective (7) is minimized when the assignment occurs in non-decreasing order, i.e. $\bar{p}_1 < \bar{p}_2 < \dots < \bar{p}_{M-1}$. To see this, and without loss of generality, suppose that the first m out of $M - 1$ terms do not obey this assignment ordering because of one or more violations. Define the following:

$$S(l, h) = \sum_{i=l}^h \prod_{j=l}^i \bar{p}_j \quad (8)$$

where l and h are positive integers and $l \leq h$. We prove that the objective value in (7) can still decrease by showing that the sum $S(1, m) = \text{ASN} - 1$ can attain a lower value if the violations are corrected. The fact that these violations occur to the first m terms $\bar{p}_1, \dots, \bar{p}_m$ is unimportant with respect to objective (7) because if the m' terms $\bar{p}_{1+s}, \dots, \bar{p}_{m'+s}$ are shuffled for some $s > 0$ then so are the first $m = m' + s$ terms $\bar{p}_1, \dots, \bar{p}_{m'+s}$. The proof of the claim is by induction on m . Sum $S(1, 1) = (\bar{p}_1)$ is minimized if \bar{p}_1 is the smallest in set $\{\bar{p}_m\}_{m=1}^{M-1}$. Assume the sum $S(1, m - 1)$ is minimized if terms $\bar{p}_1, \dots, \bar{p}_{m-1}$ are the $m - 1$ smallest elements in set $\{\bar{p}_m\}_{m=1}^{M-1}$ and $\bar{p}_1 < \dots < \bar{p}_{m-1}$. By noting that $S(1, m)$ is a sum of products of positive terms, probabilities $\bar{p}_1, \dots, \bar{p}_m$ should be the m smallest ones of set $\{\bar{p}_m\}_{m=1}^{M-1}$. In addition, $S(1, m) = S(1, m - 1) + (\bar{p}_1 * \bar{p}_2 * \dots * \bar{p}_m)$. The first operand is minimized when $\bar{p}_1 < \dots < \bar{p}_{m-1}$ by the induction step. The second operand value does not depend on the ordering of terms $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_m$. Thus $S(1, m)$ is minimized when $\bar{p}_1 < \dots < \bar{p}_m$ and the induction holds true.

Therefore, by neglecting the constraints on $\bar{p}_1, \bar{p}_2, \dots, \bar{p}_{M-1}$ and their inter-dependencies, the ASN is minimized by reducing each term \bar{p}_m and giving higher weight to the earlier ones. However, this solution is infeasible for a GSPRT because \bar{p}_m is the probability that the test does not stop at stage m given the test is at that stage. Thus the value of \bar{p}_m impacts distributions $\{f_{m'}(L_{m'} | H_1)\}$ and consequently modifies the constraints on $\{\bar{p}_{m'}\}, m' > m$ at the given specifications $P_{\mathcal{M}}$ and $P_{\mathcal{F}}$. This violates the non-decreasing order of the above solution. Still, a greedy design of the GSPRT tries to mimic this solution for $\{\bar{p}_m\}$ but considers these additional constraints. It then optimizes every product in $S(1, M - 1)$ in a sequential manner, which gives higher weights for terms \bar{p}_m that show first. This totals to $M - 1$ simple optimization problems. By neglecting the inter-dependencies, the greedy design falls behind an optimal algorithm that improves all the components of objective $S(1, M - 1)$ under all constraints and dependencies in a single complex problem.

C. Generalized Approach to GSPRT Design

Refer to the GSPRT design algorithm as Algorithm 1. This is a generalization of the greedy algorithm. While the latter only optimizes a component of objective (7) at a time starting with the early components, Algorithm 1 proceeds in the same manner but spans a wider window and achieves higher performance at the expense of solving a larger optimization problem.

Given the test is at stage m , the probability of a miss at this stage is

$$P_{\mathcal{M}_m}(a_m) = \int_{-\infty}^{a_m} f_m(L_m | H_1) dL_m \quad (9)$$

while that of a false alarm is

$$P_{\mathcal{F}_m}(b_m = -a_m) = \int_{-a_m}^{\infty} f_m(L_m | H_0) dL_m \quad (10)$$

Note that detecting H_1 at stage m is $1 - P_{\mathcal{M}_m}(-a_m)$. We have

$$P_{\mathcal{M}_m}(a_m) + \bar{p}_m - P_{\mathcal{M}_m}(-a_m) = 0 \quad (11)$$

since probabilities of all outcomes at stage m under H_1 add up to unity. Moreover, to preserve symmetry we have

$$P_{\mathcal{M}_m}(a_m) = P_{\mathcal{F}_m}(-a_m) \quad (12)$$

Define D as the *depth* of Algorithm 1. By sliding a window of size D over the individual products in $S(1, m)$, each windowed objective is optimized over D constrained and inter-dependent parameters $\{\bar{p}_m\}$. The lower d elements of the minimizer are used to compute the thresholds of the GSPRT using (11), and the window is shifted by d , with a potential overlap with its prior position. Thus D controls the complexity of the optimization sub-problems while both D and d determine their count. Algorithm 1 is detailed below. Upon finding $\{\bar{p}_m\}$, the computation of the thresholds follows easily from (11). The greedy and optimal algorithms are special cases of Algorithm 1 with parameters $(D = 1, d = 1)$ and $(D = M - 1, d = M - 1)$ respectively. Note that no single value of parameters D and d should be imposed, which allows to adapt to time-varying observation distributions.

Algorithm 1

Input:

- Depth D , increment d
- Distributions $\{f'_1(l_1|H_k), \dots, f'_{M-1}(l_{M-1}|H_k)\}$

Output:

- probabilities $\bar{p}_1^*, \dots, \bar{p}_{M-1}^*$

 1: **procedure:**

 2: **for** $m = 1$ to $(M - 1)$ at increments of d **do**

 3: Find $\{\hat{p}_{m'}\}_{m'=m}^{m+D-1} = \underset{\{\bar{p}_{m'}\}_{m'=m}^{m+D-1}}{\operatorname{argmin}} S(m, m + D - 1)$

4: s.t.

 5: $P_{\mathcal{M},m}(a_m) + \sum_{m'=m}^{m+D-2} \bar{p}_{m'} P_{\mathcal{M},m'+1}(a_{m'+1}) \leq P_{\mathcal{M}}$

 6: and (3), (11) and (12) true for all $\{\bar{p}_{m'}\}_{m'=m}^{m+D-1}$

 7: $(\bar{p}_m^*, \dots, \bar{p}_{m+d-1}^*) \leftarrow (\hat{p}_m, \dots, \hat{p}_{m+d-1})$

 8: **if** $m + D - 1 \geq M - 1$ **then**

 9: $(\bar{p}_{m+d}^*, \dots, \bar{p}_{M-1}^*) \leftarrow (\hat{p}_{m+d}, \dots, \hat{p}_{M-1})$

 10: **break**

 11: **end if**

 12: **end for**

 13: **end procedure**

V. SIMULATION RESULTS AND ANALYSIS

This section illustrates how Algorithm 1 works via simulation. Though Algorithm 1 works for general independent distributions of observations Y_n , the latter are chosen to be all identically distributed. This is because we then have a closed form for the optimal solution of the GSPRT, though this is not an absolutely exact statement since the closed-form solution exists for the case where the thresholds are time-invariant: $a_n = \log\left(\frac{P_{\mathcal{M}}}{1 - P_{\mathcal{F}}}\right)$. We only apply this constraint to the SPRT which we simulate against three other GSPRT designs: $D = 1$ (greedy), $D = 2$ and $D = 3$. Increment d is one for the three tests. Observations Y_n are Gaussian of distribution $\mathcal{N}(\mu, \sigma^2)$ under H_1 and $\mathcal{N}(0, \sigma^2)$ under H_0 . The first argument of $\mathcal{N}(\cdot, \cdot)$ is the mean and the second argument is the variance. In turn the log-likelihoods l_n are also Gaussian: $l_n \sim \mathcal{N}\left(\frac{\mu^2}{2\sigma^2}, \frac{\mu^2}{\sigma^2}\right)$ under H_1 and $l_n \sim \mathcal{N}\left(-\frac{\mu^2}{2\sigma^2}, \frac{\mu^2}{\sigma^2}\right)$ under H_0 . Let $\frac{\mu}{\sigma} = 1$, and $P_{\mathcal{F}} = P_{\mathcal{M}} = 10^{-5}$. Each sequential test is simulated under H_1 over a newly generated stream of observations for 10^7 runs, allowing for a good measure of each test statistics.

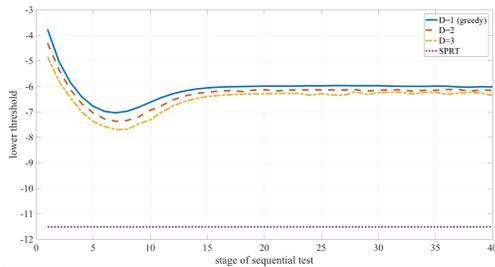


Fig. 1. Lower thresholds a_n of three GSPRTs and the SPRT. The upper thresholds b_n are their symmetric with respect to the zero axis.

Figure 1 shows the generated lower thresholds a_n of the three GSPRTs and the SPRT. While for the SPRT a_n is constant over time, this is not the case for the three other tests. Due to the difficulty of computing the distributions $g_n(L_n|H_k)$ of the bounded random variable $L_n \in (a_n, b_n)$

as well as the convolution in (3) for the Gaussian case, we apply a crude approximation to distributions $f_n(L_n|H_k)$ for $n > 1$ by fitting Gaussian distributions for the cumulative log-likelihood L_n . Due to these approximations, the constraint in line 5 of Algorithm 1 is no more tightly met though it is easy to prove that there is always a feasible GSPRT design solution upon using the exact distributions $f_n(L_n|H_k)$ since Gaussian observations Y_n are continuous and unbounded. Still, the resultant miss probabilities are $1.01 * 10^{-3}$, $8.42 * 10^{-4}$ and $6.97 * 10e - 4$ for depths $D = 1$, $D = 2$ and $D = 3$ respectively, which is fairly low. For the SPRT the error is $4.9 * 10^{-6} < P_{\mathcal{M}}$. Despite the approximations, the results show an important trend: the miss probability drops for larger depth values D . Concordantly, the thresholds of the GSPRT push towards the optimal solution for higher values of D . Note that a GSPRT with large enough D and exact distributions $f_n(L_n|H_k)$ should perform at least as the SPRT since the thresholds can still be time-varying.

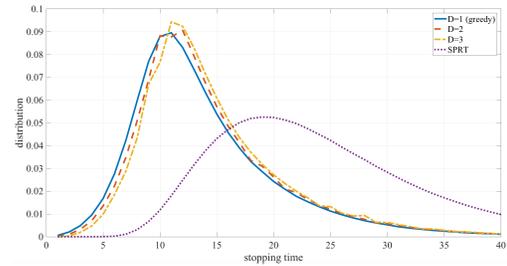


Fig. 2. Distribution of stopping time N of three GSPRTs and the SPRT.

Another important metric to look at in the evaluation of the sequential test is its stopping time. By setting an upper bound of 40 on the number of collected observations, all the tests but the SPRT terminated more than 99% of the times. The SPRT on the other hand concluded 93.1% of the times. Figure 2 shows the distribution of the stopping time for the four tests. The fat tail of the SPRT curve tells that the SPRT achieves low-error performance by collecting more observations. It is remarkable that there is minimal shift of the stopping times towards higher values as depth D of the GSPRT increases despite the improvement in the error performance. This hints that it is sufficient to consider a finite window of distributions of the observations to achieve a target performance of the sequential test. In addition, this points out the flexibility of the GSPRT in suppressing the error at a minimum ASN, which is gained by the flexibility in setting the GSPRT thresholds.

VI. CONCLUSION

In this paper a parametrized technique for designing the GSPRT thresholds is presented. With the parametrization, this becomes a generalization for greedy and optimal GSPRT design algorithms. The technique applies for general distributions of the observations assuming independence, and presents flexibility in setting the tradeoff between design complexity and error performance of the sequential test. Three GSPRTs and the SPRT are simulated. The major bottleneck is in the evaluation of the distributions of the sufficient statistic over the stages of the test. Resolving this issue should still prove the algorithm as more efficient and facilitate tweaking its design parameters.

REFERENCES

- [1] T. Lai, *Sequential Analysis: Some Classical Problems and New Challenges*. Technical report (Stanford University. Department of Statistics), Department of Statistics, Stanford University, 2000.
- [2] A. Wald, *Sequential Analysis*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics, J. Wiley & Sons, Incorporated, 1947.
- [3] V. P. Draglia, A. G. Tartakovsky, and V. V. Veeravalli, "Multi-hypothesis sequential probability ratio tests .i. asymptotic optimality," *IEEE Transactions on Information Theory*, vol. 45, pp. 2448–2461, Nov 1999.
- [4] Y. Liu and S. D. Blostein, "Optimality of the sequential probability ratio test for nonstationary observations," *IEEE Transactions on Information Theory*, vol. 38, pp. 177–182, Jan 1992.
- [5] X. Li, J. Liu, and Z. Ying, "Generalized sequential probability ratio test for separate families of hypotheses," *Sequential Analysis*, vol. 33, no. 4, pp. 539–563, 2014.
- [6] V. I. Cochlar, Ji, "On the optimum sequential test of two hypotheses for statistically dependent observations," *Kybernetika*, vol. 14, no. 1, pp. (57)–69, 1978.
- [7] R. Iyer and A. Tewfik, "Optimal ordering of observations for fast sequential detection," in *Signal Processing Conference (EU-SIPCO), 2012 Proceedings of the 20th European*, pp. 126–130, 2012.
- [8] D. P. Malladi and J. L. Speyer, "A generalized shiryayev sequential probability ratio test for change detection and isolation," *IEEE Transactions on Automatic Control*, vol. 44, pp. 1522–1534, Aug 1999.