On optimal quantization rules for some problems in sequential decentralized detection

XuanLong Nguyen[1], Martin J. Wainwright[1,2] and Michael I. Jordan[1,2]
[1] Department of Electrical Engineering and Computer Science
[2] Department of Statistics
University of California, Berkeley
{xuanlong,wainwrig,jordan}@eecs.berkeley.edu

Abstract

We consider the design of systems for sequential decentralized detection, a problem that entails several interdependent choices: the choice of a stopping rule (specifying the sample size), a global decision function (a choice between two competing hypotheses), and a set of quantization rules (the local decisions on the basis of which the global decision is made). This paper addresses the problem of whether in the Bayesian formulation of sequential decentralized detection, optimal local decision functions can be found within the class of stationary rules. We develop an asymptotic approximation to the optimal cost of stationary quantization rules and exploit this approximation to show that stationary quantizers are not optimal in general. We also consider the class of blockwise stationary quantizers, and show that asymptotically optimal quantizers are likelihood-based threshold rules.[1]

Keywords: sequential detection; decentralized detection; hypothesis testing; experimental design; quantizer design; decision-making under constraints.

I. Introduction

Detection is a classical discrimination or hypothesis-testing problem, in which observations $\{X_1, X_2, \ldots\}$ are assumed to be drawn i.i.d. from the (multivariate) conditional distribution $\mathbb{P}(\cdot \,|\, H)$ and the goal is to infer the value of the random variable H, which takes values in $\{0, 1\}$. In a typical engineering application, the case $\{H = 1\}$ represents the presence of some target to be detected, whereas $\{H = 0\}$ represents its

[1]This work was presented in part at the International Symposium on Information Theory, July 2006, Seattle, WA.

absence. Placing this problem in a communication-theoretic context, a *decentralized detection* problem is a hypothesis-testing problem in which the decision-maker is not given access to the raw data points X_n, but instead must infer H based only on the output of a set of *quantization rules* or *local decision functions*, say $\{U_n = \phi_n(X_n)\}$, which map the raw data to quantized values. This basic problem of decentralized detection has been studied extensively for several decades [17], [19], [6]; see the overview papers [20], [23], [3], [5] and references therein for more background. Of interest in this paper is the extension to an-online setting: more specifically, the *sequential decentralized detection* problem [19], [21], [12] involves a data sequence, $\{X_1, X_2, \ldots\}$, and a corresponding sequence of summary statistics, $\{U_1, U_2, \ldots\}$, determined by a sequence of local decision rules $\{\phi_1, \phi_2, \ldots\}$. The goal is to design both the local decision functions and to specify a global decision rule so as to predict H in a manner that optimally trades off accuracy and delay. In short, the sequential decentralized detection problem is the communication-constrained extension of classical formulation of sequential centralized decision-making problems (see, e.g., [8], [15], [10]) to the decentralized setting.

In setting up a general framework for studying sequential decentralized problems, Veeravalli et al. [22] defined five problems, denoted "Case A" through "Case E," distinguished from one another by the amount of information available to the local sensors. In applications such as power-constrained sensor networks, one cannot assume that the decision-maker and sensors can communicate over a high-bandwidth channel, nor that the sensors have unbounded memory. Most suited to this perspective—and the focus of this paper—is Case A, in which the local decisions are of the simplified form $\phi_n(X_n)$; i.e., neither local memory nor feedback are assumed to be available. Noting that Case A is not amenable to dynamic programming and hence presumably intractable, Veeravalli et al. [22] suggested restricting the analysis to the class of *stationary* local decision functions; i.e., local decision functions ϕ_n that are independent of n. They conjectured that stationary decision functions might actually be optimal in the setting of Case A (given the intuitive symmetry and high degree of independence of the problem in this case), even though it is not possible to verify this optimality via DP arguments. This conjecture has remained open since it was first posed by Veeravalli et al. [22], [21].

The main contribution of this paper is to resolve this question by showing that stationary decision functions are, in fact, *not* optimal for decentralized problems of type A. Our argument is based on an asymptotic characterization of the optimal Bayesian risk as the cost per sample goes to zero. In this asymptotic regime, the optimal cost can be expressed as a simple function of priors and Kullback-Leibler (KL) divergences. This characterization allows us to construct counterexamples to the stationarity conjecture, both in an exact and an asymptotic setting. In the latter setting, we present a class of problems in which there always exists a range of prior probabilities for which stationary strategies, either deterministic or randomized, are suboptimal. We note in passing that an intuition for the source of this suboptimality is easily provided—it is due to the asymmetry of the KL divergence.

It is well known that optimal quantizers when unrestricted are necessarily likelihood-based threshold rules [19]. Our counterexamples and analysis imply that optimal thresholds are not generally stationary (i.e., the threshold may differ from sample to sample). We also provide a partial converse to this result: specifically, if we restrict ourselves to stationary (or blockwise stationary) quantizer designs, then there exists an optimal design that is a threshold rule based on the likelihood ratio. We prove this result by establishing a quasiconcavity result for the asymptotically optimal cost function. In this paper, this result is proven for the space of deterministic quantizers with arbitrary output alphabets, as well as for the space of randomized quantizers with binary outputs. We conjecture that the same result holds more generally for randomized quantizers with arbitrary output alphabets.

The remainder of this paper is organized as follows. We begin in Section II with background on the Bayesian formulation of sequential detection problems, and Wald's approximation. Section III provides a simple asymptotic approximation of the optimal cost that underlies our main analysis in Section IV. In Section V, we establish the existence of optimal decision rules that are likelihood-based threshold rules, under the restriction to blockwise stationarity. We conclude with a discussion in Section VI.

II. BACKGROUND

This section provides background on the Bayesian formulation of sequential (centralized) detection problems. Of particular use in our subsequent analysis is Wald's approximation of the cost of optimal sequential test.

Let $\mathbb{P}_0$ and $\mathbb{P}_1$ represent the distribution of X, when conditioned on $\{H = 0\}$ and $\{H = 1\}$ respectively. Assume that $\mathbb{P}_0$ and $\mathbb{P}_1$ are absolutely continuous with respect to one another. We use $f^0(x)$ and $f^1(x)$ to denote the respective density functions with respect to some dominating measure (e.g., Lebesgue for continuous variables, or counting measure for discrete-valued variables).

Our focus is the Bayesian formulation of the sequential detection problem [15], [21]; accordingly, we let $\pi^1 = \mathbb{P}(H = 1)$ and $\pi^0 = \mathbb{P}(H = 0)$ denote the prior probabilities of the two hypotheses. Let $X_1, X_2, \ldots$ be a sequence of conditionally i.i.d. realizations of X. A sequential decision rule consists of a *stopping time* N defined with respect to the sigma field $\sigma(X_1, \ldots, X_N)$, and a decision function γ

measurable with respect to $\sigma(X_1, \ldots, X_N)$. The cost function is the expectation of a weighted sum of the sample size N and the probability of incorrect decision—namely

$$J(N, \gamma) := \mathbb{E}\big\{cN + \mathbb{I}[\gamma(X_1, \ldots, X_N) \neq H]\big\}, \tag{1}$$

where $c > 0$ is the incremental cost of each sample. The overall goal is to choose the pair (N, γ) so as to minimize the expected loss (1).

It is well known that the optimal solution of the sequential decision problem can be characterized recursively using dynamic programming (DP) arguments [1], [25], [15], [2]. Although useful in classical (centralized) sequential detection, the DP approach is not always straightforward to apply to *decentralized* versions of sequential detection [21]. In the remainder of this section, we describe an asymptotic approximation of the optimal sequential cost, originally due to Wald (cf. [16]), valid as $c \to 0$. To sketch out Wald's approximation, we begin by noting the optimal stopping rule for the cost function (1) takes the form

$$N = \inf\big\{n \geq 1 \;\big|\; L_n(X_1, \ldots, X_n) := \sum_{i=1}^{n} \log \frac{f^1(X_i)}{f^0(X_i)} \notin (a, b)\big\}, \tag{2}$$

for some real numbers $a < b$. Given this stopping rule, the optimal decision function has the form

$$\gamma(L_N) = \begin{cases} 1 & \text{if } L_N \geq b, \\ 0 & \text{if } L_N \leq a. \end{cases} \tag{3}$$

Consider the two types of error:

$$\begin{aligned} \alpha &= \mathbb{P}_0(\gamma(L_N) \neq H) = \mathbb{P}_0(L_N \geq b) \\ \beta &= \mathbb{P}_1(\gamma(L_N) \neq H) = \mathbb{P}_1(L_N \leq a). \end{aligned}$$

As $c \to 0$, it can be shown that the optimal choice of a and b satisfies $a \to -\infty, b \to \infty$, and the corresponding α, β satisfy $\alpha + \beta \to 0$. Ignoring the overshoot of L_N upon the optimal stopping time N (i.e., instead assuming L_N attains precisely the value a or b) we can express a, b, $\mathbb{E}N$ and the cost function J in terms of α and β as follows [24]:

$$a \approx a(\alpha, \beta) := \log \frac{\beta}{1 - \alpha} \quad \text{and} \quad b \approx b(\alpha, \beta) := \log \frac{1 - \beta}{\alpha} \tag{4}$$

$$\mathbb{E}_0[L_N] \approx (1 - \alpha)a + \alpha b \quad \text{and} \quad \mathbb{E}_1[L_N] \approx (1 - \beta)b + \beta a \tag{5}$$

Now define the Kullback-Leibler divergences

$$D^1 = \mathbb{E}_1[\log \frac{f^1(X_1)}{f^0(X_1)}] = D(f^1 || f^0), \qquad \text{and} \qquad D^0 = -\mathbb{E}_0[\log \frac{f^1(X_1)}{f^0(X_1)}] = D(f^0 || f^1). \tag{6}$$

With a slight abuse of notation, we shall also use $D(\alpha,\beta)$ to denote a function in $[0,1]^2 \to \mathbb{R}$ such that:

$$D(\alpha,\beta) := \alpha \log \frac{\alpha}{\beta} + (1-\alpha)\log\frac{1-\alpha}{1-\beta}.$$

With the above approximations, the cost function J of the decision rule based on envelopes a and b can be written as

$$\begin{aligned} J &= \pi^1 \mathbb{E}_1(cN + \mathbb{I}[L_N \leq a]) + \pi^0 \mathbb{E}_0(cN + \mathbb{I}[L_N \geq b]) \\ &= c\pi^1 \frac{\mathbb{E}_1 L_N}{D^1} + c\pi^0 \frac{\mathbb{E}_0 L_N}{-D^0} + \pi^0\alpha + \pi^1\beta, \qquad (7) \\ &\approx c\pi^0 \frac{D(\alpha, 1-\beta)}{D^0} + c\pi^1 \frac{D(1-\beta,\alpha)}{D^1} + \pi^0\alpha + \pi^1\beta, \qquad (8) \end{aligned}$$

where the third line follows from Wald's equation [24]. Let $\widetilde{J}(\alpha,\beta)$ denote the approximation (8) of J.

Let J^* denote the cost of an optimal sequential test, i.e.,

$$J^* = \inf_{a,b} J. \qquad (9)$$

A useful result due Chernoff [7] states that under certain assumption (to be elaborated in the next section), J^* has the following form:

$$J^* \approx (\frac{\pi^0}{D^0} + \frac{\pi^1}{D^1}) c \log c^{-1}(1 + o(1)). \qquad (10)$$

III. Characterization of optimal stationary quantizers

Turning now to the decentralized setting, the primary challenge lies in the design of the quantization rules ϕ_n applied to data X_n. When X_n is univariate, a deterministic quantization rule ϕ_n is a function that maps $\mathcal{X}$ to the discrete space $\mathcal{U} = \{0,\ldots,K-1\}$ for some natural number K. For multivariate X_n with d dimensions arising in the multiple sensor setting, a deterministic quantizer ϕ_n is defined as a mapping from the d-dimensional product space $\mathcal{X}$ to $\mathcal{U} = \{0,\ldots,K-1\}^d$. In the decentralized problem defined as Case A by Veeravalli et al. [22], the function ϕ_n is composed of d separate quantizer functions, one each for each dimension. A randomized quantizer ϕ_n is obtained by placing a distribution over the space of deterministic quantizers.

Any fixed set of quantization rules ϕ_n yields a sequence of compressed data $U_n = \phi_n(X_n)$, to which the classical theory can be applied. We are thus interested in choosing quantization rules $\phi_1,\phi_2,\ldots$ so that the error resulting from applying the optimal sequential test to the sequence of statistics $U_1,U_2,\ldots$ is minimized over some space Φ of quantization rules. For a given quantizer ϕ_n we use

$$f^i_{\phi_n}(u) := \mathbb{P}_i(\phi_n(X_n) = u), \qquad \text{for} \quad i = 0,1,$$

to denote the distributions of the compressed data, conditioned on the hypothesis. In general, when randomized quantizers are allowed, the vector $(f^0_{\phi_n}(.), f^1_{\phi_n}(.))$ ranges over a convex set, denoted $\operatorname{conv}\Phi$, whose extreme points correspond to deterministic quantizers based on likelihood ratio threshold rules [18].

We say that a quantizer design is *stationary* if the rule ϕ_n is independent of n; in this case, we simplify the notation to f^1_ϕ and f^0_ϕ. In addition, we define the KL divergences $D^1_\phi := D(f^1_\phi || f^0_\phi)$ and $D^0_\phi := D(f^0_\phi||f^1_\phi)$. Moreover, let J_ϕ and J^*_ϕ denote the analogues of the functions J in Eq. (7) and J^* in Eq. (9), respectively, defined using D^i_ϕ, for $i = 0, 1$. In this scenario, the sequence of compressed data $U_1, \ldots, U_n, \ldots$ are drawn i.i.d. from either f^0_ϕ or f^1_ϕ. Thus we can use the approximation (10) to characterize the asymptotically optimal stationary quantizer design. This is stated formally in the lemma to follow.

We begin by stating the assumptions underlying the lemma. For a given class of quantizers Φ, we assume that the Kullback-Leibler divergences are uniformly bounded away from zero

$$D(f^1_\phi||f^0_\phi) > 0, D(f^0_\phi||f^1_\phi) > 0 \text{ for all } \phi \in \Phi \tag{11}$$

and moreover that the variance of the log likelihood ratios are bounded

$$\sup_{\phi\in\Phi}\{\mathrm{Var}_{f^1_\phi} \log(f^1_\phi/f^0_\phi)) < \infty, \qquad \text{and} \qquad \sup_{\phi\in\Phi} \mathrm{Var}_{f^0_\phi} \log(f^1_\phi/f^0_\phi)) < \infty. \tag{12}$$

Lemma 1. *(a) Under assumptions* (11) *and* (12)*, the optimal stationary cost takes the form*

$$J^*_\phi = \left(\frac{\pi^0}{D^0_\phi} + \frac{\pi^1}{D^1_\phi}\right) c \log c^{-1}\ (1 + r_\phi) \tag{13}$$

where $|r_\phi| = o(1)$ *as* $c \to 0$.

(b) If $\sup_{\phi\in\Phi} \max\{\log(f^1_\phi/f^0_\phi), \log(f^0_\phi/f^1_\phi)\} < M$ *for some constant* M*, then* (13) *holds with* $\sup_{\phi\in\Phi} |r_\phi| = o(1)$ *as* $c \to 0$.

Proof: (a) This part is immediate from a combination of Theorems 1 and 2 of Chernoff [7].

(b) We begin by bounding the error in the approximation (8). By definition of the stopping time N, we have either (i) $b \le L_N \le b + M$ or (ii) $a - M \le L_N \le a$. By standard arguments due to Wald [24], it is simple to obtain $e^b\alpha \le 1 - \beta \le e^{b+M}\alpha$, or equivalently $b \le b(\alpha, \beta) = \log \frac{1-\beta}{\alpha} \le b + M$. Similar reasoning for case (ii) yields $a - M \le a(\alpha, \beta) = \log \frac{\beta}{1-\alpha} \le a$. Now, note that

$$\mathbb{E}_0 L_N = \alpha\mathbb{E}_0[L_N|L_N \ge b] + (1 - \alpha)\mathbb{E}_0[L_N|L_N \le a].$$

Conditioning on the event $L_N \in [b, b + M]$, we have $|L_N - b(\alpha, \beta)| \le M$. Similarly, conditioning on the event $L_N \in [a - M, a]$, we have $|L_N - b(\alpha, \beta)| \le M$. This yields $|\mathbb{E}_0 L_N - (-D(\alpha, 1 - \beta))| \le M$.

Similar reasoning yields $|\mathbb{E}_1 L_N - D(1-\beta,\alpha)| \le M$. Let $\widetilde{J}_\phi(a,b)$ denote the approximation (8) of J_ϕ. We obtain:

$$|J_\phi - \widetilde{J}_\phi(\alpha,\beta)| \le 2cM.$$

Note that the approximation error bound is independent of ϕ. Thus, it suffices to establish the asymptotic behavior (13) for the quantity $\inf_{\alpha,\beta} \widetilde{J}_\phi(\alpha,\beta)$, where the infimum is taken over pairs of realizable error probabilities (α,β). Moreover, we only need to consider the asymptotic regime $\alpha+\beta \to 0$, since the error probabilities α and β vanish as $c \to 0$. It is simple to see that $D(1-\beta,\alpha) = \log(1/\alpha)(1+o(1))$, and $D(1-\alpha,\beta) = \log(1/\beta)(1+o(1))$. Hence, $\inf_{\alpha,\beta} \widetilde{J}_\phi(\alpha,\beta)$ can be expressed as

$$\inf_{\alpha,\beta}\left\{\pi^0\alpha + \pi^1\beta + c\pi^0\frac{\log(1/\beta)}{D^0_\phi} + c\pi^1\frac{\log(1/\alpha)}{D^1_\phi}\right\}(1+o(1)). \tag{14}$$

This infimum, taken over all positive (α,β), is achieved at $\alpha^* = \frac{c\pi^1}{D^1_\phi \pi^0}$ and $\beta^* = \frac{c\pi^0}{D^0_\phi\pi^1}$. Plugging the quantities α^* and β^* into Eq. (14) yields (13). Note that the asymptotic quantity $o(1)$ in (13) is absolutely bounded by $\alpha^*+\beta^* \to 0$ uniformly for all quantizer ϕ, because D^1_ϕ and D^0_ϕ are uniformly bounded away from zero due to the Lemma's assumption.

It remains to show that error probabilities (α^*,β^*) can be approximately realized by using a sufficiently large threshold $b > 0$ and small threshold $a < 0$ while incurring an approximation cost of order $O(c)$ uniformly for all ϕ. Indeed, let us choose thresholds a' and b' such that $e^{-(b'+M)}/2 \le \alpha^* \le e^{-b'}$, and $e^{a'-M}/2 \le \beta^* \le e^{a'}$. Let α' and β' be the corresponding errors associated with these two thresholds. As before, we also have $\alpha' \in (e^{-(b'+M)}/2, e^{-b'})$ and $\beta' \in (e^{a'-M}/2, e^{a'})$. Clearly, $|\alpha^* - \alpha'| \le e^{-b'}(1 - e^{-M}/2) = O(\alpha^*) = O(c)$. Similarly, $|\beta^*-\beta'| = O(c)$. By the mean value theorem,

$$|\log(1/\alpha^*) - \log(1/\alpha')| \le |\alpha^*-\alpha'|e^{b'+M} \le 2e^M(1-e^{-M}/2) = O(1).$$

Similarly, $\log(1/\beta^*) - \log(1/\beta') = O(1)$. Hence, the approximation of (α^*,β^*) by the realizable (α',β') incurs a cost at most $O(c)$. Furthermore, the constant in the asymptotic bound $O(c)$ is independent of quantizer $\phi \in \Phi$.

□

For the rest of this paper, we shall assume that all assumptions of Lemma 1 hold.

Remarks:

1) The preceding approximation of the optimal cost essentially ignores the overshoot of the likelihood ratio L_N. While it is possible to analyze this overshoot to obtain a finer approximation (cf. [11], [16], [10], [14]), we see that this is not needed for our purpose. Lemma 1 shows that given a fixed

prior (π^0, π^1), among all stationary quantizer designs in Φ, ϕ is optimal for sufficiently small c if *and* only if ϕ minimizes what we shall call the *sequential cost coefficient*:

$$G_\phi := \frac{\pi^0}{D^0_\phi} + \frac{\pi^1}{D^1_\phi}.$$

2) As a consequence of Lemma 7 to be proved in the sequel, if we consider the class Φ of all binary randomized quantizers, then sequential cost coefficient G_ϕ is a quasiconcave function with respect to $(f^0_\phi(.), f^1_\phi(.))$. (A function F is quasiconcave if and only if for any η, the level set $\{F(x) \geq \eta\}$ is a convex set; see Boyd and Vandenberghe [4] for further background). The minimum of a quasiconcave function lies in the set of extreme points in its domain. For the set $\operatorname{conv}\Phi$, these extreme points can be realized by deterministic quantizers based on likelihood ratios [20]. Consequently, we conclude that for quantizers with binary outputs, the optimal cost is not decreased by considering randomized quantizers. We conjecture that this statement also holds beyond the binary case.

Section V is devoted to a more detailed study of asymptotically optimal stationary quantizers. In the meantime, we turn to the question of whether stationary quantizers are optimal in either finite-sample or asymptotic settings.

IV. Suboptimality of stationary designs

It was shown by Tsitsiklis [19] that optimal quantizers ϕ_n take the form of threshold rules based on the likelihood ratio $f^1(X_n)/f^0(X_n)$. Veeravalli et al. [22], [21] asked whether these rules can always be taken to be stationary, a conjecture that has remained open. In this section, we resolve this question with a negative answer in both the finite-sample and asymptotic settings.

A. Suboptimality in exact setting

We begin by providing a numerical counterexample for which stationary designs are suboptimal. Consider a problem in which $X \in \mathcal{X} = \{1, 2, 3\}$ and the conditional distributions take the form

$$f^0(x) = \begin{bmatrix} \frac{8}{10} & \frac{1999}{10000} & \frac{1}{10000} \end{bmatrix} \text{ and } f^1(x) = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Suppose that the prior probabilities are $\pi^1 = \frac{8}{100}$ and $\pi^0 = \frac{92}{100}$, and that the cost for each sample is $c = \frac{1}{100}$.

If we restrict to binary quantizers (i.e., $\mathcal{U} = \{0, 1\}$), by the symmetric roles of the output alphabets there are only three possible deterministic quantizers:

1) Design A: $\phi_A(X_n) = 0 \iff X_n = 1$. As a result, the corresponding distribution for U_n is specified by $f^0_{\phi_A}(u_n) = [\frac{4}{5}\ \frac{1}{5}]$ and $f^1_{\phi_A}(u) = [\frac{1}{3}\ \frac{2}{3}]$.
2) Design B: $\phi_B(X_n) = 0 \iff X_n \in \{1, 2\}$. The corresponding distribution for U_n is given by $f^0_{\phi_B}(u) = [\frac{9999}{10000}\ \frac{1}{10000}]$ and $f^1_{\phi_B}(u) = [\frac{2}{3}\ \frac{1}{3}]$.
3) Design C: $\phi_C(X_n) = 0 \iff X_n \in \{1, 3\}$. The corresponding distribution for U_n is specified by $f^0_{\phi_C} \sim [\frac{8001}{10000}\ \frac{1999}{10000}]$ and $f^1_{\phi_C}(u) = [\frac{2}{3}\ \frac{1}{3}]$.

Now consider the three stationary strategies, each of which uses only one fixed design, A, B or C. For any given stationary quantization rule ϕ, we have a classical centralized sequential problem, for which the optimal cost (achieved by a sequential probability ratio test) can be computed using a dynamic-programming procedure [25], [1]. Accordingly, for each stationary strategy, we compute the optimal cost function J for 10^6 points on the p-axis by performing 300 updates of Bellman's equation (cf. [2]). In all cases, the difference in cost between the 299th and 300th updates is less than 10^{-6}. Let J_A, J_B and J_C denote the optimal cost function for sequential tests using all A's, all B's, and all C's, respectively. When evaluated at $\pi^1 = 0.08$, these computations yield $J_A = 0.0567, J_B = 0.0532$ and $J_C = 0.08$.

Finally, we consider a non-stationary rule obtained by applying design A for only the first sample, and applying design B for the remaining samples. Again using Bellman's equation, we find that the cost for this design is

$$J_* = \min\{\min\{\pi^1, 1-\pi^1\}, c + J_B(P(H=1|u_1=0))P(u_1=0) + \\ J_B(P(H=1|u_1=1))P(u_1=1)\} = 0.052767,$$

which is better than any of the stationary strategies.

In this particular example, the cost J^* of the non-stationary quantizer yields a slim improvement (0.0004) over the best stationary rule J_B. This slim margin is due in part to the choice of a small per-sample cost $c = 0.01$; however, larger values of c do not yield counterexample when using the particular distributions specified above. A more significant factor is that our non-stationary rule differs from the optimal stationary rule B only in its treatment of the first sample. This fact suggests that one might achieve better cost by alternating between using design A and design B on the odd and even samples, respectively. Our analysis of the asymptotic setting in the next section confirms this intuition.

B. *Asymptotic suboptimality for both deterministic and randomized quantizers*

We now prove that in a broad class of examples, there is a range of prior probabilities for which stationary quantizer designs are suboptimal. Our result stems from the following observation: Lemma 1 implies that in order to achieve a small cost we need to choose a quantizer ϕ for which the KL divergences $D^0_\phi := D(f^0_\phi || f^1_\phi)$ and $D^1_\phi := D(f^1_\phi || f^0_\phi)$ are both as large as possible. Due to the asymmetry of the KL divergence, however, these maxima are not necessarily achieved by a single quantizer ϕ. This suggests that one could improve upon stationary designs by applying different quantizers to different samples, as the following lemma shows.

Lemma 2. *Let ϕ_1 and ϕ_2 be any two quantizers. If the following inequalities hold*

$$D^0_{\phi_1} < D^0_{\phi_2} \text{ and } D^1_{\phi_1} > D^1_{\phi_2} \tag{15}$$

then there exists a non-empty interval $(U, V) \subseteq (0, +\infty)$ such that as $c \to 0$,

$$\begin{aligned} J^*_{\phi_1} \leq J^*_{\phi_1,\phi_2} \leq J^*_{\phi_2} \quad &\text{if} \quad \frac{\pi^0}{\pi^1} \leq U \\ J^*_{\phi_1,\phi_2} < \min\{J^*_{\phi_1}, J^*_{\phi_2}\} - \Theta(c \log c^{-1}) \quad &\text{if} \quad \frac{\pi^0}{\pi^1} \in (U, V) \\ J^*_{\phi_1} \geq J^*_{\phi_1,\phi_2} \geq J^*_{\phi_2} \quad &\text{if} \quad \frac{\pi^0}{\pi^1} \geq V, \end{aligned}$$

where $J^*_{\phi_1,\phi_2}$ denotes the optimal cost of a sequential test that alternates between using ϕ_1 and ϕ_2 on odd and even samples respectively.

Proof: According to Lemma 1, we have

$$J^*_{\phi_i} = \left(\frac{\pi^0}{D^0_{\phi_i}} + \frac{\pi^1}{D^1_{\phi_i}} \right) c \log c^{-1} (1 + o(1)), \quad i = 0, 1. \tag{16}$$

Now consider the sequential test that applies quantizers ϕ_1 and ϕ_2 alternately to odd and even samples. Furthermore, let this test consider two samples at a time. Let $f^0_{\phi_1\phi_2}$ and $f^1_{\phi_1\phi_2}$ denote the induced conditional probability distributions, jointly on the odd-even pairs of quantized variables. From the additivity of the KL divergence and assumption (15), there holds:

$$D(f^0_{\phi_1\phi_2} || f^1_{\phi_1\phi_2}) = D^0_{\phi_1} + D^0_{\phi_2} > 2D^0_{\phi_1} \tag{17a}$$

$$D(f^1_{\phi_1\phi_2} || f^0_{\phi_1\phi_2}) = D^1_{\phi_1} + D^1_{\phi_2} < 2D^1_{\phi_1}. \tag{17b}$$

Clearly, the cost of the proposed sequential test is an upper bound for $J^*_{\phi_1,\phi_2}$. Furthermore, the gap between this upper bound and the true optimal cost is no more than $O(c)$. Hence, as in the proof of

Lemma 1, as $c \to 0$, the optimal cost $J^*_{\phi_1,\phi_2}$ can be written as

$$\left(\frac{2\pi^0}{D^0_{\phi_1} + D^0_{\phi_2}} + \frac{2\pi^1}{D^1_{\phi_1} + D^1_{\phi_2}}\right) c \log c^{-1}(1 + o(1)). \tag{18}$$

From equations (16) and (18), simple calculations yield the claim with

$$U = \frac{D^0_{\phi_1}(D^1_{\phi_1} - D^1_{\phi_2})(D^0_{\phi_1} + D^0_{\phi_2})}{D^1_{\phi_1}(D^1_{\phi_1} + D^1_{\phi_2})(D^0_{\phi_2} - D^0_{\phi_1})} < V = \frac{D^0_{\phi_2}(D^1_{\phi_1} - D^1_{\phi_2})(D^0_{\phi_1} + D^0_{\phi_2})}{D^1_{\phi_2}(D^1_{\phi_1} + D^1_{\phi_2})(D^0_{\phi_2} - D^0_{\phi_1})}. \tag{19}$$

□

Example: Let us return to the example provided in the previous subsection. Note that the two quantizers ϕ_A and ϕ_B satisfy assumption (15), since $D(f^0_{\phi_B}||f^1_{\phi_B}) = 0.4045 < D(f^0_{\phi_A}||f^1_{\phi_A}) = 0.45$ and $D(f^1_{\phi_B}||f^0_{\phi_B}) = 2.4337 > D(f^1_{\phi_A}||f^0_{\phi_A}) = 0.5108$. Furthermore, both quantizers dominates ϕ_C in terms of KL divergences: $D(f^0_{\phi_C}||f^1_{\phi_C}) = 0.0438$, $D(f^0_{\phi_C}||f^1_{\phi_C}) = 0.0488$. As a result, there exist a range of priors for which a sequential test using stationary quantizer design (either ϕ_A, ϕ_B or ϕ_C for all samples) is not optimal.

Theorem 3. *(a) Suppose that Φ is a finite collection of quantizers, and that there is no single quantizer ϕ that dominates all other quantizers in Φ in the sense that*

$$D^0_{\phi} \;\geq\; D^0_{\phi'} \quad \textit{and} \quad D^1_{\phi} \;\geq\; D^1_{\phi'} \qquad \textit{for all} \quad \phi' \in \Phi. \tag{20}$$

Then there exists a non-empty range of prior probabilities for which no stationary design based on a quantizer in Φ is optimal.

(b) For any non-deterministic ϕ in the randomized class $\operatorname{conv}\Phi$, *there exists a non-stationary quantizer design that has strictly smaller sequential cost coefficient than that of a stationary design based on ϕ for any choice of prior probabilities.*

Proof: (a) Since there are a finite number of quantizers in Φ and no quantizer dominates all others, the interval $(0,\infty)$ is divided into at least two adjacent non-empty intervals, each of which corresponds to a range of prior probability ratios π^0/π^1 for which a quantizer is strictly optimal (asymptotically) among all stationary designs. Let them be (δ_1, δ) and (δ, δ_2), for two quantizers, namely, ϕ_1 and ϕ_2. In particular, δ is the value for π^0/π^1 for which the sequential cost coefficients are equal—viz. $G_{\phi_1} = G_{\phi_2}$—which happens only if assumption (15) holds. Some calculations verify that

$$\delta = \frac{D^0_{\phi_1} D^0_{\phi_2}(D^1_{\phi_2} - D^1_{\phi_1})}{D^1_{\phi_1} D^1_{\phi_2}(D^0_{\phi_1} - D^0_{\phi_2})}. \tag{21}$$

By Lemma 2, a non-stationary design obtained by alternating between ϕ_1 and ϕ_2 has smaller sequential cost than both ϕ_1 and ϕ_2 for $\pi^0/\pi^1 \in (U, V)$, where U and V are given in equation (19). Since it can be verified that δ as defined (21) belongs to the interval (U, V), we conclude that for $\pi^0/\pi^1 \in (U, V) \cap (\delta_1, \delta_2)$, this non-stationary design has smaller cost than any stationary design using $\phi \in \Phi$.

(b) Let $\phi \in \operatorname{conv} \Phi$ be a randomized quantizer (i.e., at each step choose with non-zero probabilities $w_1, \ldots, w_k$ from quantizers $\phi_1, \ldots, \phi_k \in \Phi$, respectively, where $\sum_{i=1}^k w_i = 1$). Clearly, the density induced by ϕ satisfy: $f^0_\phi = \sum_{i=1}^k w_i f^0_{\phi_i}$ and $f^1_\phi = \sum_{i=1}^k w_i f^1_{\phi_i}$. Due to strict convexity of the KL divergence functional with respect jointly to the two density arguments [9], by Jensen's inequality we have: $D^0_\phi < \sum_{i=1}^k w_i D^0_{\phi_i}$ and $D^1_\phi < \sum_{i=1}^k w_i D^1_{\phi_i}$. Since $D^0_{\phi_i}$ and $D^1_{\phi_i}$ are bounded from above uniformly for all $\phi_i \in \Phi$, it is possible to approximate $(w_1, \ldots, w_k)$ by rational numbers of the form $(q_1/N, q_2/N, \ldots, q_k/N)$ for some natural numbers $q_1, \ldots, q_k$ and N satisfying $\sum_{i=1}^k q_i = N$ such that

$$\begin{aligned} D^0_\phi &< \sum_{i=1}^k q_i D^0_{\phi_i}/N \\ D^1_\phi &< \sum_{i=1}^k q_i D^1_{\phi_i}/N. \end{aligned}$$

Now consider the non-stationary quantizer that applies ϕ_1 for q_1 steps, then ϕ_2 for q_2 steps and so on, up to ϕ_k for q_k steps, yielding a total of N steps, and then repeats this sequence starting again at step $N+1$. By construction, this non-stationary quantizer has a smaller cost than that of quantizer ϕ for any choice of prior. $\square$

Remarks: It is interesting to contrast the Bayesian formulation of the problem of quantizer design with the Neyman-Pearson formulation. Our results on the suboptimality of stationary quantizer design in the Bayesian formulation repose on the asymmetry of the Kullback-Leibler divergence, as well as the sensitivity of the optimal quantizers on the prior probability. We note that Mei [12] (see p. 58) considered the Neyman-Pearson formulation of this problem. In this formulation, it can be shown that for all sequential tests for which the Type 1 and Type 2 errors are bounded by α and β, respectively, then as $\alpha + \beta \to 0$, the expected stopping time $\mathbb{E}_0 N$ under hypothesis $H = 0$ is asymptotically minimized by applying a stationary quantizer ϕ^* that maximizes $D(f^0_\phi || f^1_\phi)$. Similarly, the expected stopping time $\mathbb{E}_1 N$ under hypothesis $H = 1$ is asymptotically minimized by the stationary quantizer ϕ^{**} that maximizes $D(f^1_\phi || f^0_\phi)$ [12]. In this context, the example in subsection IV-A provides a case in which the asymptotically minimal KL divergences ϕ^* and ϕ^{**} are not the same, due to the asymmetry, which suggests that there may not exist a stationary quantizer that simultaneously minimizes both $\mathbb{E}_1 N$ and

$\mathbb{E}_0 N$.

C. Asymptotic suboptimality in multiple sensor setting

Our analysis thus far has established that with a single sensor per time step ($d = 1$), applying multiple quantizers to different samples can reduce the sequential cost. As pointed out by one of the referees, it is natural to ask whether the same phenomenon persists in the case of multiple sensors ($d > 1$). In this section, we show that the phenomenon does indeed carry over, more specifically by providing an example in which stationary strategies are still sub-optimal in comparison to non-stationary ones. The key insight is that we have only a fixed number of dimensions, whereas as $c \to 0$ we are allowed to take more samples, and each sample can act as an extra dimension, providing more flexibility for non-stationary strategies.

Suppose that the observation vector X_n at time n is d-dimensional, with each component corresponding to a sensor in a typical decentralized setting. Suppose that the observations from each sensor are assumed to be independent and identically distributed according to the conditional distributions defined in our earlier example (see Section IV-A). Of interest are the optimal deterministic binary quantizer designs for all d sensors. Although there are three possible choices ϕ_A, ϕ_B and ϕ_C for each sensor, the quantizer ϕ_C is dominated by the other two, so each sensor should choose either ϕ_A and ϕ_B. Suppose that among these sensors, a subset of size k choose ϕ_A and whereas the remaining $d-k$ sensors choose ϕ_B for $0 \leq k \leq d$. We thus have $d+1$ possible stationary designs to consider. For each k, the sequential cost coefficient corresponding to the associated stationary design takes the form

$$G_k := \frac{\pi^0}{kD^0_{\phi_A} + (d-k)D^0_{\phi_B}} + \frac{\pi^1}{kD^1_{\phi_A} + (d-k)D^1_{\phi_B}}. \tag{22}$$

Now consider the following non-stationary design: the first sensor alternates between decision rules ϕ_A and ϕ_B, while the remaining $d-1$ sensors simply apply the stationary design based on ϕ_B. For this design, the associated sequential cost coefficient is given by

$$G := \frac{2\pi^0}{D^0_{\phi_A} + (2d-1)D^0_{\phi_B}} + \frac{2\pi^1}{D^1_{\phi_A} + (2d-1)D^1_{\phi_B}}. \tag{23}$$

Consider the interval (U, V), where the interval has endpoints

$$U = \frac{D^1_{\phi_B} - D^1_{\phi_A}}{D^0_{\phi_A} - D^0_{\phi_B}} \frac{D^0_{\phi_A} + (2d-1)D^0_{\phi_B}}{D^1_{\phi_A} + (2d-1)D^1_{\phi_B}} \frac{D^0_{\phi_B}}{D^1_{\phi_B}} < V = \frac{D^1_{\phi_B} - D^1_{\phi_A}}{D^0_{\phi_A} - D^0_{\phi_B}} \frac{D^0_{\phi_A} + (2d-1)D^0_{\phi_B}}{D^1_{\phi_A} + (2d-1)D^1_{\phi_B}} \frac{D^0_{\phi_A} + (d-1)D^0_{\phi_B}}{D^1_{\phi_A} + (d-1)D^1_{\phi_B}}.$$

Straightforward calculations yield that for any prior likelihood $\pi^0/\pi^1 \in (U, V)$, the minimal cost over stationary designs $\min_{k=0,\ldots,d} G_k$ is strictly larger than the sequential cost G of the non-stationary design, previously defined in equation (23).

V. On asymptotically optimal blockwise stationary designs

Despite the possible loss in optimality, it is useful to consider some form of stationarity in order to reduce computational complexity of the optimization and decision process. In this section, we consider the class of *blockwise stationary* designs, meaning that there exists some natural number T such that $\phi_{T+1} = \phi_1, \phi_{T+2} = \phi_2$, and so on. For each T, let C_T denote the class of all blockwise stationary designs with period T. We assume throughout the analysis that each decision rule ϕ_n $(n = 1, \ldots, T)$ satisfies conditions (11) and (12). Thus, as T increases, we have a hierarchy of increasingly rich quantizer classes that will be seen to yield progressively better approximations to the optimal solution.

For a fixed prior (π^0, π^1) and $T > 0$, let $(\phi_1, \ldots, \phi_T)$ denote a quantizer design in C_T. As before, the cost J^*_ϕ of an asymptotically optimal sequential test using this quantizer design is of order $c \log c^{-1}$ with the sequential cost coefficient

$$G_\phi = \frac{T\pi^0}{D^0_{\phi_1} + \ldots + D^0_{\phi_T}} + \frac{T\pi^1}{D^1_{\phi_1} + \ldots + D^1_{\phi_T}}. \tag{24}$$

G_ϕ is a function of the vector of probabilities introduced by the quantizer: $(f^0_\phi(.), f^1_\phi(.))$. We are interested in the properties of a quantization rule ϕ that minimizes J^*_ϕ.

It is well known that there exist optimal quantizers—*when unrestricted*— that can be expressed as threshold rules based on the log likelihood ratio (LLR) [19]. Our counterexamples in the previous sections imply that the thresholds need not be stationary (i.e., the threshold may differ from sample to sample). In the remainder of this section, we addresses a partial converse to this issue: specifically, if we restrict ourselves to stationary (or blockwise stationary) quantizer designs, then there exists an optimal design consisting of LLR-based threshold rules.

It turns out that the analysis for the case $T > 1$ can be reduced to an analysis that is closely related to our earlier analysis for $T = 1$. Indeed, consider the sequential cost coefficient for the time step $n = 1$, where the rules for the other time steps are held fixed. From (24) we have

$$G_\phi = \frac{T\pi^0}{D^0_{\phi_1} + s_0} + \frac{T\pi^1}{D^1_{\phi_1} + s_1},$$

for non-negative constants s_0 and s_1. As we will show, our earlier analysis of the sequential cost coefficient, in which $s_0 = s_1 = 0$, carries through to the case in which these values are non-zero. This allows us to provide (in Theorem 9) a characterization of the optimal blockwise stationary quantizer.

Definition 4. *The quantizer design function* $\phi : \mathcal{X} \rightarrow \mathcal{U}$ *is said to be a* likelihood ratio threshold rule *if there are thresholds* $d_0 = -\infty < d_1 < \ldots < d_K = +\infty$*, and a permutation* $(u_1, \ldots, u_K)$ *of* $(0, 1, \ldots, K-1)$ *such that for* $l = 1, \ldots, K$*, with* $\mathbb{P}_0$*-probability 1, we have:*

$$\phi(X) = u_l \text{ if } d_{l-1} \leq f^1(X)/f^0(X) \leq d_l,$$

When $f^1(X)/f^0(X) = d_{l-1}$*, set* $\phi(X) = u_{l-1}$ *or* $\phi(X) = u_l$ *with* $\mathbb{P}_0$*-probability 1.*[2]

Previous work on the extremal properties of likelihood ratio based quantizers guarantees that the Kullback-Leibler divergence is maximized by a LLR-based quantizer [18]. In our case, however, the sequential cost coefficient G_ϕ involves a pair of KL divergences, D^0_ϕ and D^1_ϕ, which are related to one another in a nontrivial manner. Hence, establishing asymptotic optimality of LLR-based rules for this cost function does not follow from existing results, but rather requires further understanding of the interplay between these two KL divergences.

The following lemma concerns certain "unnormalized" variants of the Kullback-Leibler (KL) divergence. Given vectors $a = (a_0, a_1)$ and $b = (b_0, b_1)$, we define functions $\tilde{D}^0$ and $\tilde{D}^1$ mapping from $\mathbb{R}^4_+$ to the real line as follows:

$$\tilde{D}^0(a, b) \quad := \quad a_0 \log \frac{a_0}{a_1} + b_0 \log \frac{b_0}{b_1} \tag{25a}$$

$$\tilde{D}^1(a, b) \quad := \quad a_1 \log \frac{a_1}{a_0} + b_1 \log \frac{b_1}{b_0}. \tag{25b}$$

These functions are related to the standard (normalized) KL divergence via the relations $\tilde{D}^0(a, 1-a) \equiv D(a_0, a_1)$, and $\tilde{D}^1(a, 1-a) \equiv D(a_1, a_0)$.

Lemma 5. *For any positive scalars* $a_1, b_1, c_1, a_0, b_0, c_0$ *such that* $\frac{a_1}{a_0} < \frac{b_1}{b_0} < \frac{c_1}{c_0}$, at least one *of the two*

[2]This last requirement of the definition is termed the *canonical* likelihood ratio quantizer by Tsitsiklis [18]. Although one could consider performing additional randomization when there are ties, our later results (in particular, Lemma 7) establish that in this case, randomization will not further decrease the optimal cost J^*_ϕ.

following conditions must hold:

$$\tilde{D}^0(a, b+c) > \tilde{D}^0(b, c+a) \quad \textit{and} \quad \tilde{D}^1(a, b+c) > \tilde{D}^0(b, c+a), \quad \textit{or} \tag{26a}$$

$$\tilde{D}^0(c, a+b) > \tilde{D}^0(b, c+a) \quad \textit{and} \quad \tilde{D}^1(c, a+b) > \tilde{D}^0(b, c+a). \tag{26b}$$

This lemma implies that under certain conditions on the ordering of the probability ratios, one can increase *both* KL divergences by re-quantizing. This insight is used in the following lemma to establish that the optimal quantizer ϕ behaves almost like a likelihood ratio rule. To state the result, recall that the *essential supremum* is the infimum of the set of all η such that $f(x) \leq \eta$ for $\mathbb{P}_0$-almost all x in the domain, for a measurable function f.

Lemma 6. *If ϕ is an asymptotically optimal quantizer, then for all pairs $(u_1, u_2) \in \mathcal{U}$, $u_1 \neq u_2$, there holds:*

$$\frac{f^1(u_1)}{f^0(u_1)} \notin \left(\operatorname{ess} \inf_{x:\phi(x)=u_2} \frac{f^1(x)}{f^0(x)}, \ \operatorname{ess} \sup_{x:\phi(x)=u_2} \frac{f^1(x)}{f^0(x)} \right).$$

Note that a likelihood ratio rule guarantees something stronger: For $\mathbb{P}_0$-almost all x such that $\phi(x) = u_1$, $f^1(x)/f^0(x)$ takes a value either to the left or to the right, but not to both sides, of the interval specified above.

Lemma 7 stated below essentially guarantees quasiconcavity of G_ϕ for the case of binary quantizers. To state the result, let $F : [0,1]^2 \to R$ be given by

$$F(a_0, a_1) = \frac{c_0}{D(a_0, a_1) + d_0} + \frac{c_1}{D(a_1, a_0) + d_1}. \tag{27}$$

Lemma 7. *For any non-negative constants c_0, c_1, d_0, d_1, the function F defined in* (27) *is quasiconcave.*

We provide a proof of this result in the Appendix. An immediate consequence of Lemma 7 is that LLR-based quantizers exist for the class of randomized quantizers with binary outputs.

Corollary 8. *Restricting to the class of (blockwise) stationary binary quantizers, there exists an asymptotically optimal quantizer ϕ that is a (deterministic) likelihood ratio threshold rule.*

Proof: Let ϕ is a (randomized) binary quantizer. The sequential cost coefficient can be written as $G_\phi = F(f^0_\phi(0), f^1_\phi(0))$. The set of $\{(f^0_\phi(0), f^1_\phi(0)\}$ for all ϕ is a convex set whose extreme points can be realized by deterministic likelihood ratio threshold rules (Prop. 3.2 of [18]). Since the minimum of a quasiconcave function must lie at one such extreme point [4], the corollary is immediate as a consequence of Lemma 7. □

It turns out that the same statement can also be proved for deterministic quantizers with arbitrary output alphabets:

Theorem 9. *Restricting to the class of (blockwise) stationary and deterministic decision rules, then there exists an asymptotically optimal quantizer ϕ that is a likelihood ratio threshold rule.*

We present the full proof of this theorem in the Appendix. The proof exploits both Lemma 6 and Lemma 7.

VI. Discussion

In this paper, we have studied the problem of sequential decentralized detection. For quantizers with neither local memory nor feedback (Case A in the taxonomy of Veeravalli et al. [22]), we have established that stationary designs need not be optimal in general. Moreover, we have shown that in the asymptotic setting (i.e., when the cost per sample goes to zero), there is a class of problems for which there exists a range of prior probabilities over which stationary strategies are suboptimal.

There are a number of open questions raised by the analysis in this paper. First, our analysis has established only that the best stationary rule chosen from a finite set of deterministic quantizers need not be optimal. Is there a corresponding example with an infinite number of deterministic stationary quantizer designs for which none is optimal? Second, Corollary 8 establishes the optimality of likelihood ratio rules for randomized decision rules that produce binary outputs. This proof was based on the quasiconcavity of the function G_ϕ that specifies the asymptotic sequential cost coefficient. Is this function G_ϕ also quasiconcave for quantizers other than binary ones? Such quasiconcavity would extend the validity of Theorem 9 for the general class of randomized quantizers.

Acknowledgements

The research described in this paper was supported in part by National Science Foundations Grants 0412995 (MJ) and DMS-0605165 (MW). We thank the reviewers for their comments and suggestions which helped to improve the manuscript.

Appendix

Proof of Lemma 5

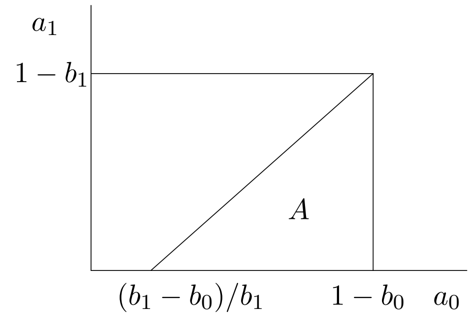

Fig. 1: Illustration of the domain A.

By renormalizing, we can assume w.l.o.g. that $a_1+b_1+c_1=a_0+b_0+c_0=1$. Also w.l.o.g, assume that $b_1 \geq b_0$. Thus, $c_1 > c_0$ and $a_1 < a_0$. Replacing $c_1 = 1-a_1-b_1$ and $c_0 = 1-a_0-b_0$, the inequality $c_1/c_0 > b_1/b_0$ is equivalent to $a_1 < a_0 b_1/b_0 - (b_1-b_0)/b_0$.

We fix values of b, and consider varying $a \in A$, where A denotes the domain for (a_0, a_1) governed by the following equality and inequality constraints: $0 < a_1 < 1-b_1;\ 0 < a_0 < 1-b_0;\ a_1 < a_0$ and

$$a_1 < a_0 b_1/b_0 - (b_1-b_0)/b_0. \tag{28}$$

Note that the third constraint ($a_1 < a_0$) is redundant due to the other three constraints. In particular, constraint (28) corresponds to a line passing through $((b_1-b_0)/b_1, 0)$ and $(1-b_0, 1-b_1)$ in the (a_0, a_1) coordinates. As a result, A is the interior of the triangle defined by this line and two other lines given by $a_1 = 0$ and $a_0 = 1-b_0$ (see Figure 1).

Since both $\tilde{D}^0(a, 1-a)$ and $\tilde{D}^1(a, 1-a)$ correspond to KL divergences, they are convex functions with respect to (a_0, a_1). In addition, the derivatives with respect to a_1 are $\frac{a_1-a_0}{a_1(1-a_1)} < 0$ and $\log\frac{a_1(1-a_0)}{a_0(1-a_1)} < 0$, respectively. Hence, both functions can be (strictly) bounded from below by increasing a_1 while keeping a_0 unchanged, i.e., by replacing a_1 by a_1' so that (a_0, a_1') lies on the line given by (28), which is equivalent to the constraint $c_1/c_0 = b_1/b_0$. Let $c_1' = 1-b_1-a_1'$; then $c_1'/c_0 = b_1/b_0$. Our argument has established inequalities (a) and (b) in the following chain of inequalities:

$$\tilde{D}^1(a, b+c) \overset{(a)}{>} a_1' \log \frac{a_1'}{a_0} + (b_1 + c_1') \log \frac{b_1 + c_1'}{b_0 + c_0} \tag{29a}$$

$$\overset{(b)}{=} a_1' \log \frac{a_1'}{a_0} + c_1' \log \frac{c_1'}{c_0} + b_1 \log \frac{b_1}{b_0} \tag{29b}$$

$$\overset{(c)}{\geq} (a_1' + c_1') \log \frac{a_1' + c_1'}{a_0 + c_0} + b_1 \log \frac{b_1}{b_0} \tag{29c}$$

$$= \tilde{D}^1(a+c, b), \tag{29d}$$

inequality (c) follows from an application of the log-sum inequality [9]. A similar conclusion holds for $\tilde{D}^0(a, b+c)$.

Proof of Lemma 6

Suppose the opposite is true, that there exist two sets S_1, S_2 with positive $\mathbb{P}_0$-measure such that $\phi(X) = u_2$ for any $X \in S_1 \cup S_2$, and

$$\frac{f^1(S_1)}{f^0(S_1)} < \frac{f^1(u_1)}{f^0(u_1)} < \frac{f^1(S_2)}{f^0(S_2)}. \tag{30}$$

By reassigning S_1 or S_2 to the quantile u_1, we are guaranteed to have a new quantizer ϕ' such that $D^0_{\phi'} > D^0_{\phi^*}$ and $D^1_{\phi'} > D^1_{\phi^*}$, thanks to Lemma 5. As a result, ϕ' has a smaller sequential cost $J^*_{\phi'}$, which is a contradiction.

Proof of Lemma 7

The proof of this lemma is conceptually straightforward, but the algebra is involved. To simplify the notation, we replace a_0 by x, a_1 by y, the function $D(a_0, a_1)$ by $f(x, y)$, and the function $D(a_1, a_0)$ by $g(x, y)$. Finally, we assume that $d_0 = d_1 = 0$; the proof will reveal that this case is sufficient to establish the more general result with arbitrary non-negative scalars d_0 and d_1.

We have $f(x, y) = x \log(x/y) + (1-x) \log[(1-x)/(1-y)]$ and $g(x, y) = y \log(y/x) + (1-y) \log[(1-y)/(1-x)]$. Note that both f and g are convex functions and are non-negative in their domains, and moreover that we have $F(x, y) = c_0/f(x, y) + c_1/g(x, y)$. In order to establish the quasiconcavity of F, it suffices to show that for any (x, y) in the domain of F, for any vector $h = [h_0 \ h_1] \in \mathbb{R}^2$ such that $h^T \nabla F(x, y) = 0$, there holds

$$h^T \nabla^2 F(x, y) \ h < 0 \tag{31}$$

(see Boyd and Vandenberghe [4]). Here we adopt the standard notation of ∇F for the gradient vector of F, and $\nabla^2 F$ for its Hessian matrix. We also use F_x to denote the partial derivative with respect to variable x, F_{xy} to denote the partial derivative with respect to x and y, and so on.

We have $\nabla F = -\frac{c_0 \nabla f}{f^2} - \frac{c_1 \nabla g}{g^2}$. Thus, it suffices to prove relation (31) for vectors of the form

$$h = \left[\left(-\frac{c_0 f_y}{f^2} - \frac{c_1 g_y}{g^2} \right) \quad \left(\frac{c_0 f_x}{f^2} + \frac{c_1 g_x}{g^2} \right) \right]^T.$$

It is convenient to write $h = c_0 v_0 + c_1 v_1$, where $v_0 = [-f_y/f^2 \quad f_x/f^2]^T$ and $v_1 = [-g_y/g^2 \quad g_x/g^2]^T$.

The Hessian matrix $\nabla^2 F$ can be written as $\nabla^2 F = c_0 H_0 + c_0 H_1$, where

$$H_0 = -\frac{1}{f^3} \begin{bmatrix} f_{xx} f - 2 f_x^2 & f_{xy} f - 2 f_x f_y \\ f_{xy} f - 2 f_x f_y & f_{yy} f - 2 f_y^2 \end{bmatrix},$$

and

$$H_1 = -\frac{1}{g^3} \begin{bmatrix} g_{xx} g - 2 g_x^2 & g_{xy} g - 2 g_x g_y \\ g_{xy} g - 2 g_x g_y & g_{yy} g - 2 g_y^2 \end{bmatrix}.$$

Now observe that

$$h^T \nabla^2 F h = (c_0 v_0 + c_1 v_1)^T (c_0 H_0 + c_1 H_1)(c_0 v_0 + c_1 v_1),$$

which can be simplified to

$$h^T \nabla^2 F h = c_0^3 v_0^T H_0 v_0 + c_1^3 v_1^T H_1 v_1 + c_0^2 c_1 (2 v_0^T H_0 v_1 + v_0^T H_1 v_0) + c_0 c_1^2 (2 v_0^T H_1 v_1 + v_1^T H_0 v_1).$$

This function is a polynomial in c_0 and c_1, which are restricted to be non-negative scalars (at least one of which is assumed to be non-zero). Therefore, it suffices to prove that all the coefficients of this polynomial (with respect to c_0 and c_1) are strictly negative. In particular, we shall show that

(i) $v_0^T H_0 v_0 \leq 0$, and

(ii) $2 v_0^T H_0 v_1 + v_0^T H_1 v_0 \leq 0$,.

where in both cases equality occurs only if $x = y$, which is outside of the domain of F. The strict negativity of the other two coefficients follows from entirely analogous arguments.

First, some straightforward algebra shows that inequality (i) is equivalent to the relation

$$f_{xx} f_y^2 + f_{yy} f_x^2 \geq 2 f_x f_y f_{xy}.$$

But note that f is a convex function, so $f_{xx} f_{yy} \geq f_{xy}^2$. Hence, we have

$$f_{xx} f_y^2 + f_{yy} f_x^2 \overset{(a)}{\geq} 2\sqrt{f_{xx} f_{yy}} |f_x f_y| \overset{(b)}{\geq} 2 f_x f_y f_{xy},$$

thereby proving (i). (In this argument, inequality (a) follows from the fact that $a^2 + b^2 \geq 2ab$, whereas inequality (b) follows from the strict convexity of f. Equality occurs only if $x = y$.)

Regarding (ii), some further algebra reduces it to the inequality

$$G_1 + G_2 - G_3 \geq 0, \tag{32}$$

where

$$\begin{aligned} G_1 &= 2(f_y g_y f_{xx} + f_x g_x f_{yy} - (f_y g_x + f_x g_y) f_{xy}), \\ G_2 &= f_y^2 g_{xx} + f_x^2 g_{yy} - 2 f_x f_y g_{xy}, \\ G_3 &= \frac{2}{g}(f_y g_x - f_x g_y)^2. \end{aligned}$$

At this point in the proof, we need to exploit specific information about the functions f and g, which are defined in terms of KL divergences. To simplify notation, we let $u = x/y$ and $v = (1-x)/(1-y)$. Computing derivatives, we have

$$\begin{aligned} f_x(x,y) &= \log(x/y) - \log((1-x)/(1-y)) = \log(u/v), \\ f_y(x,y) &= (1-x)/(1-y) - x/y = v - u, \\ g_x(x,y) &= (1-y)/(1-x) - y/x = 1/v - 1/u, \\ g_y(x,y) &= \log(y/x) - \log((1-y)/(1-x)) = \log(v/u), \\ \nabla^2 f(x,y) &= \begin{bmatrix} \frac{1}{x(1-x)} & -\frac{1}{y(1-y)} \\ -\frac{1}{x(1-x)} & \frac{1-x}{(1-y)^2} + \frac{x}{y^2} \end{bmatrix}, \qquad \text{and} \qquad \nabla^2 g(x,y) = \begin{bmatrix} \frac{1-y}{(1-x)^2} + \frac{y}{x^2} & -\frac{1}{x(1-x)} \\ -\frac{1}{x(1-x)} & \frac{1}{y(1-y)} \end{bmatrix}. \end{aligned}$$

Noting that $f_x = -g_y$; $g_{xy} = -f_{xx}$; $f_{xy} = -g_{yy}$, we see that equation (32) is equivalent to

$$2(f_x g_x f_{yy} + f_y g_x g_{yy}) - f_x^2 g_{yy} + f_y^2 g_{xx} \;\; \geq \;\; \frac{2}{g}(f_y g_x - f_x g_y)^2. \tag{33}$$

To simplify the algebra further, we shall make use of the inequality $(\log t^2)^2 \leq (t - 1/t)^2$, which is valid for any t. This implies that

$$f_y g_x = (v-u)(1/v - 1/u) \leq f_x g_y = -(\log(u/v))^2 = -f_x^2 = -g_y^2 \leq 0.$$

Thus, $-f_x^2 g_{yy} \geq f_y g_x g_{yy}$, and $\frac{2}{g}(f_y g_x - f_x g_y)^2 \leq \frac{2}{g} f_y g_x (f_y g_x - f_x g_y)$. As a result, (33) would follow if we can show that

$$2(f_x g_x f_{yy} + f_y g_x g_{yy}) + f_y g_x g_{yy} + f_y^2 g_{xx} \geq \frac{2}{g} f_y g_x (f_y g_x - f_x g_y).$$

For all $x \neq y$, we may divide both sides by $-f_y(x,y)g_x(x,y) > 0$. Consequently, it suffices to show that:

$$-2f_x f_{yy}/f_y - f_y g_{xx}/g_x - 3g_{yy} \geq \frac{2}{g}(f_x g_y - g_x f_y),$$

or, equivalently,

$$2\log(u/v)\left(\frac{v}{u-1} + \frac{u}{1-v}\right) + \left(\frac{u}{1-x} + \frac{v}{x}\right) - \frac{3}{y(1-y)} \geq \frac{2}{g}\left(\frac{(u-v)^2}{uv} - (\log\frac{u}{v})^2\right),$$

or, equivalently,

$$2\log(u/v)\frac{(u-v)(u+v-1)}{(u-1)(1-v)} + \frac{(u-v)^2(u+v-4uv)}{uv(u-1)(1-v)} \geq \frac{2}{g}\left(\frac{(u-v)^2}{uv} - (\log\frac{u}{v})^2\right). \quad (34)$$

Due to the symmetry, it suffices to prove (34) for $x < y$. In particular, we shall use the following inequality for logarithm mean [13], which holds for $u \neq v$:

$$\frac{3}{2\sqrt{uv} + (u+v)/2} < \frac{\log u - \log v}{u - v} < \frac{1}{(uv(u+v)/2)^{1/3}}.$$

We shall replace $\frac{\log(u/v)}{u-v}$ in (34) by appropriate upper and lower bounds. In addition, we shall also bound $g(x,y)$ from below, using the following argument. When $x < y$, we have $u < 1 < v$, and

$$\begin{aligned} g(x,y) &= y\log\frac{y}{x} + (1-y)\log\frac{1-y}{1-x} > \frac{3y(y-x)}{2\sqrt{xy} + (x+y)/2} + \frac{(1-y)(x-y)}{[(1-x)(1-y)(1-(x+y)/2)]^{1/3}} \\ &= \frac{3(1-v)(1-u)}{(u-v)(2\sqrt{u} + \frac{u+1}{2})} + \frac{(u-1)(1-v)}{(u-v)(v(v+1)/2)^{1/3}} > 0. \end{aligned}$$

Let us denote this lower bound by $q(u,v)$.

Having got rid of the logarithm terms, (34) will hold if we can prove the following:

$$\frac{6(u-v)^2(u+v-1)}{(2\sqrt{uv} + (u+v)/2)(u-1)(1-v)} + \frac{(u-v)^2(u+v-4uv)}{uv(u-1)(1-v)} \geq \frac{2}{q(u,v)}\left(\frac{(u-v)^2}{uv} - \frac{9(u-v)^2}{(2\sqrt{uv} + (u+v)/2)^2}\right),$$

or equivalently,

$$\begin{aligned} &\left(\frac{6(u+v-1)}{(2\sqrt{uv} + (u+v)/2)} + \frac{(u+v-4uv)}{uv}\right)\left(\frac{3}{(v-u)(2\sqrt{u} + \frac{u+1}{2})} - \frac{1}{(v-u)(v(v+1)/2)^{1/3}}\right) \\ &\qquad \geq 2\left(\frac{1}{uv} - \frac{9}{(2\sqrt{uv} + (u+v)/2)^2}\right), \end{aligned} \quad (35)$$

which is equivalent to

$$\begin{aligned} &\frac{(u+v-2\sqrt{uv})((u+v)/2 + 3\sqrt{uv} + 4uv)}{(2\sqrt{uv} + (u+v)/2)uv} \frac{3(v(v+1)/2)^{1/3} - (2\sqrt{u} + (u+1)/2)}{(v-u)(2\sqrt{u} + (u+1)/2)(v(v+1)/2)^{1/3}} \\ &\qquad \geq \frac{(u+v-2\sqrt{uv})((u+v)/2 + 5\sqrt{uv})}{uv(2\sqrt{uv} + (u+v)/2)^2} \end{aligned} \quad (36)$$

and also equivalent to

$$((u+v)/2+2\sqrt{uv})((u+v)/2+3\sqrt{uv}+4uv)[3(v(v+1)/2)^{1/3}-(2\sqrt{u}+(u+1)/2)]$$
$$\geq (2\sqrt{u}+(u+1)/2)(v(v+1)/2)^{1/3}((u+v)/2+5\sqrt{uv})(v-u). \quad (37)$$

It can be checked by tedious but straightforward calculus that inequality (37) holds for any $u \leq 1 \leq v$, and equality holds when $u = 1 = v$, i.e., $x = y$.

Proof of Theorem 9

Suppose that ϕ is not a likelihood ratio rule. Then there exist positive $\mathbb{P}_0$-probability disjoint sets S_1, S_2, S_3 such that for any $X_1 \in S_1, X_2 \in S_2, X_3 \in S_3$,

$$\phi(X_1) = \phi(X_3) = u_1 \quad (38a)$$

$$\phi(X_2) = u_2 \neq u_1 \quad (38b)$$

$$\frac{f^1(X_1)}{f^0(X_1)} < \frac{f^1(X_2)}{f^0(X_2)} < \frac{f^1(X_3)}{f^0(X_3)}. \quad (38c)$$

Define the probability of the quantiles as:

$$f^0(u_1) := \mathbb{P}_0(\phi(X) = u_1), \quad \text{and} \quad f^0(u_2) := \mathbb{P}_0(\phi(X) = u_2),$$
$$f^1(u_1) := \mathbb{P}_1(\phi(X) = u_1), \quad \text{and} \quad f^1(u_2) := \mathbb{P}_1(\phi(X) = u_2).$$

Similarly, for the sets S_1, S_2 and S_3, we define

$$a_0 = f^0(S_1), \quad b_0 = f^0(S_2) \quad \text{and} \quad c_0 = f^0(S_3),$$
$$a_1 = f^1(S_1), \quad b_1 = f^1(S_2), \quad \text{and} \quad c_1 = f^1(S_3).$$

Finally, let p_0, p_1, q_0 and q_1 denote the probability measures of the "residuals":

$$p_0 \;=\; f^0(u_2) - b_0, \qquad p_1 \;=\; f^1(u_2) - b_1,$$
$$q_0 \;=\; f^0(u_1) - a_0 - c_0, \qquad q_1 \;=\; f^1(u_1) - a_1 - c_1.$$

Note that we have $\frac{a_1}{a_0} < \frac{b_1}{b_0} < \frac{c_1}{c_0}$. In addition, the sets S_1 and S_3 were chosen so that $\frac{a_1}{a_0} \leq \frac{q_1}{q_0} \leq \frac{c_1}{c_0}$. From Lemma 6, there holds $\frac{p_1+b_1}{p_0+b_0} = \frac{f^1(u_2)}{f^0(u_2)} \notin \left(\frac{a_1}{a_0}, \frac{c_1}{c_0}\right)$. We may assume without loss of generality that $\frac{p_1+b_1}{p_0+b_0} \leq \frac{a_1}{a_0}$. Then, $\frac{p_1+b_1}{p_0+b_0} < \frac{b_1}{b_0}$, so $\frac{p_1}{p_0} < \frac{p_1+b_1}{p_0+b_0}$. Overall, we are guaranteed to have the ordering

$$\frac{p_1}{p_0} < \frac{p_1+b_1}{p_0+b_0} \leq \frac{a_1}{a_0} < \frac{b_1}{b_0} < \frac{c_1}{c_0}. \quad (39)$$

Our strategy will be to modify the quantizer ϕ only for those X for which $\phi(X)$ takes the values u_1 or u_2, such that the resulting quantizer is defined by a LLR-based threshold, and has a smaller (or equal) value of the corresponding cost J^*_ϕ. For simplicity in notation, we use $\mathcal{A}$ to denote the set with measures under $\mathbb{P}_0$ and $\mathbb{P}_1$ equal to a_0 and a_1; the sets $\mathcal{B}$, $\mathcal{C}$, $\mathcal{P}$ and $\mathcal{Q}$ are defined in an analogous manner. We begin by observing that we have either $\frac{a_1}{a_0} \leq \frac{q_1+a_1}{q_0+a_0} < \frac{b_1}{b_0}$ or $\frac{b_1}{b_0} < \frac{q_1+c_1}{q_0+c_0} \leq \frac{c_1}{c_0}$. Thus, in our subsequent manipulation of sets, we always bundle $\mathcal{Q}$ with either $\mathcal{A}$ or $\mathcal{C}$ accordingly without changing the ordering of the probability ratios. Without loss of generality, then, we may disregard the corresponding residual set corresponding to $\mathcal{Q}$ in the analysis to follow.

In the remainder of the proof, we shall show that either one of the following two modifications of the quantizer ϕ will improve (decrease) the sequential cost J^*_ϕ:

(i) Assign $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ to the same quantization level u_1, and leave $\mathcal{P}$ to the level u_2, or

(ii) Assign $\mathcal{P}$, $\mathcal{A}$ and $\mathcal{B}$ to the same level u_2, and leave c to the level u_1.

It is clear that this modified quantizer design respects the likelihood ratio rule for the quantization indices u_1 and u_2. By repeated application of this modification for every such pair, we are guaranteed to arrive at a likelihood ratio quantizer that is optimal, thereby completing the proof.

Let a'_0, b'_0, c'_0, p'_0 be normalized versions of a_0, b_0, c_0, p_0, respectively (i.e., $a'_0 = a_0/(p_0+a_0+b_0+c_0)$, and so on). Similarly, let a'_1, b'_1, c'_1, p'_1 be normalized versions of a_1, b_1, c_1, p_1, respectively. With this notation, we have the relations

$$
\begin{aligned}
D^0_\phi &= \sum_{u \neq u_1, u_2} f^0(u) \log \frac{f^0(u)}{f^1(u)} + (p_0 + b_0) \log \frac{p_0 + b_0}{p_1 + b_1} + (a_0 + c_0) \log \frac{a_0 + c_0}{a_1 + c_1} \\
&= A_0 + (f^0(u_1) + f^0(u_2)) \left((p'_0 + b'_0) \log \frac{p'_0 + b'_0}{p'_1 + b'_1} + (a'_0 + c'_0) \log \frac{a'_0 + c'_0}{a'_1 + c'_1} \right) \\
&= A_0 + (f^0(u_1) + f^0(u_2)) \tilde{D}^0(p' + b', a' + c'), \\
D^1_\phi &= \sum_{u \neq u_1, u_2} f^1(u) \log \frac{f^1(u)}{f^0(u)} + (p_1 + b_1) \log \frac{p_1 + b_1}{p_0 + b_0} + (a_1 + c_1) \log \frac{a_1 + c_1}{a_0 + c_0} \\
&= A_1 + (f^1(u_1) + f^1(u_2)) \tilde{D}^1(p' + b', a' + c'),
\end{aligned}
$$

where we define

$$\begin{aligned} A_0 &:= \sum_{u\neq u_1,u_2} f^0(u)\log\frac{f^0(u)}{f^1(u)} + (f^0(u_1)+f^0(u_2))\log\frac{f^0(u_1)+f^0(u_2)}{f^1(u_1)+f^1(u_2)} \geq 0, \\ A_1 &:= \sum_{u\neq u_1,u_2} f^1(u)\log\frac{f^1(u)}{f^0(u)} + (f^1(u_1)+f^1(u_2))\log\frac{f^1(u_1)+f^1(u_2)}{f^0(u_1)+f^0(u_2)} \geq 0 \end{aligned}$$

due to the non-negativity of the KL divergences.

Note that from (39) we have

$$\frac{p_1'}{p_0'} < \frac{p_1'+b_1'}{p_0'+b_0'} \leq \frac{a_1'}{a_0'} < \frac{b_1'}{b_0'} < \frac{c_1'}{c_0'},$$

in addition to the normalization constraints that $p_0' + a_0' + b_0' + c_0' = p_1' + a_1' + b_1' + c_1' = 1$. It follows that $\frac{p_1'+b_1'}{p_0'+b_0'} < \frac{p_1'+a_1'+b_1'+c_1'}{p_0'+a_0'+b_0'+c_0'} = 1$.

Let us consider varying the values of a_1', b_1', while fixing all other variables and ensuring that all the above constraints hold. Then, $a_1' + b_1'$ is constant, and both $\tilde{D}^0(p'+b', a'+c')$ and $\tilde{D}^1(p'+b', a'+c')$ increase as b_1' decreases and a_1' increases. In other words, if we define $a_0'' = a_0'$, $b_0'' = b_0'$ and a_1'' and b_1'' such that

$$\frac{a_1''}{a_0'} = \frac{b_1''}{b_0'} = \frac{1-p_1'-c_1'}{1-p_0'-c_0'},$$

then we have

$$\tilde{D}^0(p'+b', a'+c') \leq \tilde{D}^0(p'+b'', a''+c') \text{ and } \tilde{D}^1(p'+b', a'+c') \leq \tilde{D}^1(p'+b'', p''+c'). \tag{40}$$

Now note that vector (b_0'', b_1'') in $\mathbb{R}^2$ is a convex combination of $(0,0)$ and $(a_0''+b_0'', a_1''+b_1'')$. It follows that $(p_0'+b_0'', p_1'+b_1'')$ is a convex combination of (p_0', p_1') and $(p_0'+a_0''+b_0'', p_1'+a_1''+b_1'') = (p_0'+a_0'+b_0', p_1'+a_1'+b_1')$.

By (40), we obtain:

$$\begin{aligned} G_\phi &= \frac{\pi^0}{A_0 + (f^0(u_1)+f^0(u_2))\tilde{D}^0(p'+b', a'+c')} + \frac{\pi^1}{A_1 + (f^1(u_1)+f^1(u_2))\tilde{D}^1(p'+b', a'+c')} \\ &\geq \frac{\pi^0}{A_0 + (f^0(u_1)+f^0(u_2))\tilde{D}^0(p'+b'', a''+c')} + \frac{\pi^1}{A_1 + (f^1(u_1)+f^1(u_2))\tilde{D}^1(p'+b'', a''+c')} \\ &= \frac{\pi^0}{A_0 + (f^0(u_1)+f^0(u_2))D(p_0'+b_0'', p_1'+b_1'')} + \frac{\pi^1}{A_1 + (f^1(u_1)+f^1(u_2))D(p_1'+b_1'', p_0'+b_0'')} \end{aligned}$$

Applying the quasiconcavity result in Lemma 7:

$$\begin{aligned} G_\phi \quad \geq \quad & \min\Bigg\{ \frac{\pi^0}{A_0 + (f^0(u_1) + f^0(u_2))D(p_0', p_1')} + \frac{\pi^1}{A_1 + (f^1(u_1) + f^1(u_2))D(p_1', p_0')}, \\ & \frac{\pi^0}{A_0 + (f^0(u_1) + f^0(u_2))D(p_0' + a_0' + b_0', p_1' + a_1' + b_1')} + \\ & \frac{\pi^1}{A_1 + (f^1(u_1) + f^1(u_2))D(p_1' + a_1' + b_1', p_0' + a_0' + b_0')} \Bigg\}. \end{aligned}$$

But the two arguments of the minimum are the sequential cost coefficient corresponding to the two possible modifications of ϕ. Hence, the proof is complete.

REFERENCES

[1] K. J. Arrow, D. Blackwell, and M. A. Girshick. Bayes and minimax solutions of sequential decision problems. *Econometrica*, 17(3/4):213–244, 1949.

[2] D. P. Bertsekas. *Dynamic Programming and Stochastic Control*, volume 1. Athena Scientific, Belmont, MA, 1995.

[3] R. S. Blum, S. A. Kassam, and H. V. Poor. Distributed detection with multiple sensors: Part ii—advanced topics. *Proceedings of the IEEE*, 85:64–79, January 1997.

[4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[5] B. Chen, L. Tong, and P. K. Varshney. Channel aware distributed detection in wireless sensor networks. *IEEE Signal Processing Magazine*, 23(4):16–26, July 2006.

[6] B. Chen and P. Willett. On the optimality of the likelihood ratio test for local sensor decision rules in the presence of non-ideal channels. *IEEE Trans. Info. Theory*, 51(2):693–699, February 2005.

[7] H. Chernoff. Sequential design of experiments. *Annals of Statistics*, 30(3):755–770, 1959.

[8] H. Chernoff. *Sequential Analysis and Optimal Design*. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM Press, 1972.

[9] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley, 1991.

[10] T. L. Lai. Sequential analysis: Some classical problems and new challenges (with discussion). *Statist. Sinica*, 11:303–408, 2001.

[11] G. Lorden. On excess over the boundary. *Annals of Statistics*, 41(2):520–527, 1970.

[12] Y. Mei. *Asymptotically optimal methods for sequential change-point detection*. PhD thesis, California Institute of Technology, 2003.

[13] D. S. Mitrinović, J. E. Pecarić, and A. M. Fink. *Classical and New Inqualities in Analysis*. Kluwer Academic Publishers, 1993.

[14] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, NY, 1994.

[15] A. N. Shiryayev. *Optimal Stopping Rules*. Springer-Verlag, 1978.

[16] D. Siegmund. *Sequential Analysis*. Springer-Verlag, 1985.

[17] R. R. Tenney and N. R. Jr. Sandell. Detection with distributed sensors. *IEEE Trans. Aero. Electron. Sys.*, 17:501–510, 1981.

[18] J. Tsitsiklis. Extremal properties of likelihood-ratio quantizers. *IEEE Trans. on Communication*, 41(4):550–558, 1993.

[19] J. N. Tsitsiklis. On threshold rules in decentralized detection. In *Proc. 25th IEEE Conf. Decision Control*, pages 232–236, 1986.

[20] J. N. Tsitsiklis. Decentralized detection. In *Advances in Statistical Signal Processing*, pages 297–344. JAI Press, 1993.

[21] V. V. Veeravalli. Sequential decision fusion: theory and applications. *Journal of the Franklin Institute*, 336:301–322, 1999.

[22] V. V. Veeravalli, T. Basar, and H. V. Poor. Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Trans. Info. Theory*, 39(2):433–442, 1993.

[23] R. Viswanathan and P. K. Varshney. Distributed detection with multiple sensors: Part i—fundamentals. *Proceedings of the IEEE*, 85:54–63, January 1997.

[24] A. Wald. *Sequential Analysis*. John Wiley and Sons, Inc., New York, 1947.

[25] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *Annals of Statistics*, 19:326–339, 1948.