

ACTIVE SEQUENTIAL HYPOTHESIS TESTING¹

BY MOHAMMAD NAGHSHVAR AND TARA JAVIDI

Qualcomm Inc. and University of California, San Diego

Consider a decision maker who is responsible to dynamically collect observations so as to enhance his information about an underlying phenomena of interest in a speedy manner while accounting for the penalty of wrong declaration. Due to the sequential nature of the problem, the decision maker relies on his current information state to adaptively select the most “informative” sensing action among the available ones.

In this paper, using results in dynamic programming, lower bounds for the optimal total cost are established. The lower bounds characterize the fundamental limits on the maximum achievable information acquisition rate and the optimal reliability. Moreover, upper bounds are obtained via an analysis of two heuristic policies for dynamic selection of actions. It is shown that the first proposed heuristic achieves asymptotic optimality, where the notion of asymptotic optimality, due to Chernoff, implies that the relative difference between the total cost achieved by the proposed policy and the optimal total cost approaches zero as the penalty of wrong declaration (hence the number of collected samples) increases. The second heuristic is shown to achieve asymptotic optimality only in a limited setting such as the problem of a noisy dynamic search. However, by considering the dependency on the number of hypotheses, under a technical condition, this second heuristic is shown to achieve a nonzero information acquisition rate, establishing a lower bound for the maximum achievable rate and error exponent. In the case of a noisy dynamic search with size-independent noise, the obtained nonzero rate and error exponent are shown to be maximum.

1. Introduction. This paper considers a generalization of the classical sequential hypothesis testing problem due to Wald [58]. Suppose there are M hypotheses among which only one is true. A Bayesian decision maker is responsible to enhance his information about the correct hypothesis in a speedy and sequential manner while accounting for the penalty of wrong declaration. In contrast to the classical sequential M -ary hypothesis testing problem [2, 22, 39], our decision maker can choose one of K available actions and, hence, exert some control over the collected samples’ “information content.” We refer to this generalization, originally tackled by Chernoff [17], as the *active* sequential hypothesis testing problem.

Received April 2013.

¹Supported in part by the industrial sponsors of UCSD Center for Wireless Communication (CWC) and Center for Networked Systems (CNS), and NSF Grants CCF-1018722 and AST-1247995.

MSC2010 subject classifications. Primary 62F03; secondary 62B10, 62B15, 62L05.

Key words and phrases. Active hypothesis testing, sequential analysis, optimal stopping, dynamic programming, feedback gain, error exponent, information acquisition rate.

The active sequential hypothesis testing problem naturally arises in a broad spectrum of applications such as medical diagnosis [6], cognition [54], sensor management [27], underwater inspection [28], generalized search [48], group testing [16] and channel coding with perfect feedback [12]. It is intuitive that at any time instant, an optimized Bayesian decision maker relies on his current belief to adaptively select the most “informative” sensing action, that is, an action that provides the highest amount of “information.” Making this intuition precise is the topic of our study.

The most well-known instance of our problem is the case of binary hypothesis testing with passive sensing ($M = 2$, $K = 1$), first studied by Wald [58]. In this instance of the problem, the optimal action at any given time is provided by a sequential probability ratio test (SPRT). There are numerous studies on the generalizations to $M > 2$ ($K = 1$) and the performance of known simple and practical heuristic tests such as MSPRT [2, 22, 39]. The generalization to the active testing case was considered by Chernoff in [17] where a heuristic randomized policy was proposed and whose asymptotic performance was analyzed. More specifically, under a certain technical assumption on uniformly distinguishable hypotheses, the proposed heuristic policy is shown to achieve asymptotic optimality where the notion of asymptotic optimality [17] denotes the relative tightness of the performance upper bound associated with the proposed policy and the lower bound associated with the optimal policy.

The problem of active hypothesis testing also generalizes another classic problem in the literature: the comparison of experiments first introduced by Blackwell [9]. This is a single-shot version of the active hypothesis testing problem in which the decision maker can choose one of several (usually two) actions/experiments to collect a single observation sample before making the final decision. There have been extensive studies [9, 21, 24, 35–37, 57] on comparing the actions. Applying various results from [9, 21] in our context of active hypothesis testing and utilizing a dynamic programming interpretation, a notion of optimal information utility, that is, an optimal measure to quantify the information gained by different sensing actions, can be derived [43]. Inspired by this view of the problem, which coincides with that promoted by DeGroot [20], we provide a set of (uniform) lower bounds for the optimal information utility. Furthermore, we provide two heuristic policies whose performance is investigated via nonasymptotic and asymptotic analysis. The first policy is shown to be asymptotically optimal, matching the performance of the scheme proposed in [17] (and follow-up works [8, 11]), and provides a benchmark for comparison when considering Chernoff’s asymptotic regime. In contrast, our second proposed policy is only shown to be asymptotically optimal in a limited setup, including that of noisy dynamic search. However, this policy has a provable advantage for large M over those proposed in the literature. More specifically, this policy can provide, under a technical condition, reliability and speedy declaration simultaneously. In information theoretic terms, this policy can be shown to achieve nonzero information

acquisition rate and, hence, to generalize Burnashev's [12] variable-length channel coding scheme. We elaborate on a complete literature survey in Section 2.2.

The remainder of this paper is organized as follows. In Section 2 we formulate the active sequential hypothesis testing problem and discuss the related works. Section 3 provides a dynamic programming formulation and characterizes a notion of optimal information utility. In Section 4 we provide three lower bounds and two upper bounds on the optimal information utility. The bounds are nonasymptotic and complementary for various values of the parameters of the problem. Section 5 states the asymptotic consequence of the bounds obtained in Section 4. In particular, the obtained bounds are used to establish notions of order and asymptotic optimality for the proposed policies (generalizing that of [17]); and characterize lower and upper bounds on the maximum achievable information acquisition rate and the optimal reliability. In Section 6 we investigate an important special case of the active hypothesis testing, namely, the noisy dynamic search. In Section 7 we discuss the technical assumptions made in our work and contrast them with the (weaker) assumptions in the literature. More specifically, we show that our first technical assumption weakens significantly one of the assumptions made in [17]. On the other hand, our second technical assumption is significantly stronger than the corresponding assumptions in the literature. However, we show that while this assumption is critical in obtaining the nonasymptotic lower and upper bounds of Section 4, it has no bearing on our asymptotic results in Section 5. Finally, we conclude the paper and discuss future work in Section 8. In the interest of brevity, we have chosen to focus our analysis, provided in the Appendix, on Theorems 1–3, whose results, to the best of our knowledge, are entirely new and whose proofs require a substantially different approach than those commonly available in the literature. In contrast, the proofs of Propositions 1–4 as well as Corollaries 1, 3, 5–7 follow similar lines of argument to the proofs in the literature or in those obtained in the Appendix and are included in the form of a supplemental article [44].

Notation: Let $[x]^+ = \max\{x, 0\}$. The indicator function $\mathbf{1}_{\{A\}}$ takes the value 1 whenever event A occurs, and 0 otherwise. For any set \mathcal{S} , $|\mathcal{S}|$ denotes the cardinality of \mathcal{S} . All logarithms are in base 2. The entropy function on a vector $\boldsymbol{\rho} = [\rho_1, \rho_2, \dots, \rho_M] \in [0, 1]^M$ is defined as $H(\boldsymbol{\rho}) = \sum_{i=1}^M \rho_i \log \frac{1}{\rho_i}$, with the convention that $0 \log \frac{1}{0} = 0$. Finally, the Kullback–Leibler (KL) divergence between two probability density functions $q(\cdot)$ and $q'(\cdot)$ on space \mathcal{Z} is defined as $D(q\|q') = \int_{\mathcal{Z}} q(z) \log \frac{q(z)}{q'(z)} dz$, with the convention $0 \log \frac{a}{0} = 0$ and $b \log \frac{b}{0} = \infty$ for $a, b \in [0, 1]$ with $b \neq 0$.

2. Problem setup and summary of the results. In Section 2.1 we formulate the problem of active sequential hypothesis testing, referred to as Problem (P) hereafter. Section 2.2 states the main contributions of the paper and provides a summary of related works.

2.1. *Problem formulation.* Here, we provide a precise formulation of our problem.

PROBLEM (P) (Active sequential hypothesis testing). Let $\Omega_M = \{1, 2, \dots, M\}$. Let $H_i, i \in \Omega_M$, denote M hypotheses of interest among which only one holds true. Let θ be the random variable that takes the value $\theta = i$ on the event that H_i is true for $i \in \Omega_M$. We consider a Bayesian scenario with prior $\rho(0) = [\rho_1(0), \rho_2(0), \dots, \rho_M(0)]$, that is, initially $P(\theta = i) = \rho_i(0) > 0$ for all $i \in \Omega_M$. \mathcal{A}_M is the set of all sensing actions which may depend on M and is assumed to be finite with $|\mathcal{A}_M| < \infty$. Let $\mathbb{P}(\mathcal{A}_M)$ denote the collection of all probability distributions on elements of \mathcal{A}_M , that is, $\mathbb{P}(\mathcal{A}_M) = \{\lambda \in [0, 1]^{|\mathcal{A}_M|} : \sum_{a \in \mathcal{A}_M} \lambda_a = 1\}$. \mathcal{Z} is the *observation space*. For all $a \in \mathcal{A}_M$, the observation kernel $q_i^a(\cdot)$ (on \mathcal{Z}) is the probability density function for observation Z when action a is taken and H_i is true. We assume that observation kernels $\{q_i^a(\cdot)\}_{i \in \Omega_M, a \in \mathcal{A}_M}$ are known and the observations are conditionally independent over time. Let L denote the penalty (loss) for a wrong declaration, that is, the penalty of selecting $H_j, j \neq i$, when H_i is true.² Let τ be the stopping time at which the decision maker retires. The objective is to find a sequence of sensing actions $A(0), A(1), \dots, A(\tau - 1)$,³ a stopping time τ and a declaration rule $d : \mathcal{A}_M^\tau \times \mathcal{Z}^\tau \rightarrow \Omega_M$ that collectively minimize the expected total cost

$$(1) \quad \mathbb{E}[\tau] + L \text{Pe},$$

where $\text{Pe} = P(d(A_0^{\tau-1}, Z_0^{\tau-1}) \neq \theta)$ denotes the probability of making a wrong declaration, and the expectation is taken with respect to the initial prior distribution $\rho(0)$ on θ as well as the distributions of action sequence, observation sequence and the stopping time.

2.2. *Overview of the results and summary of the related works.* The first attempt to solve Problem (P) goes back to Chernoff’s work on active binary composite hypothesis testing [17]. Chernoff proposed the following scheme to select actions: at each time t , find the most likely true hypothesis, and then select an action that can discriminate this hypothesis the best from each and every element in the set corresponding to the alternative hypothesis. Much of the subsequent literature extended this approach [1, 8, 11, 31, 32, 34, 47]. Chernoff showed that as L goes to infinity, the relative difference between the expected total cost achieved by his proposed scheme and the optimal expected total cost approaches zero, which

²In general, we can define a loss matrix $[L_{ij}]_{i, j \in \Omega_M}$, where L_{ij} denotes the penalty (loss) of selecting H_j when H_i is true.

³We assume that $A(t)$ is selected as a (possibly randomized) function of $A_0^{t-1} := [A(0), A(1), \dots, A(t - 1)]$ and $Z_0^{t-1} := [Z(0), Z(1), \dots, Z(t - 1)]$, that is, sensing actions and observations up to time t .

he termed as *asymptotic optimality*.⁴ One of the main drawbacks of Chernoff's asymptotic optimality notion was his neglecting the complementary role of asymptotic analysis in M . In particular, the notion of asymptotic optimality in L falls short in showing the tension between using an (asymptotically) large number of samples to discriminate among a few hypotheses with (asymptotically) high accuracy or an (asymptotically) large number of hypotheses with a lower degree of accuracy. As a result, although the scheme proposed in [17] and its subsequent extensions [1, 8, 11, 31, 32, 34, 47] are asymptotically optimal in L , their provable information acquisition rate is restricted to zero. Intuitively, the rate of information acquisition under any given heuristic relates to the ratio between $\log M$ and the expected number of samples: the larger this ratio, the faster information is acquired.

As elaborated in Section 5.3, to obtain asymptotic characterization of the optimal expected total cost in a nonzero rate regime, it is important to propose schemes which scale optimally with M as well. In his seminal paper [12], Burnashev tackled the primal (constrained) version of Problem (P) in the context of channel coding with feedback, and provided lower and upper bounds on the expected number of samples (or, equivalently, channel uses) required to convey one of M uniformly distributed messages over a discrete memoryless channel (DMC) with a desired probability of error. The lower bound identified the dominating terms in both number of messages and error probability, hence characterized the optimal reliability function (also known as the error exponent) in addition to the feedback capacity (which was known to coincide with the Shannon capacity [53]). In this paper, we generalize⁵ this lower bound to the problem of active sequential hypothesis testing, that is, Problem (P):

- We derive three lower bounds on the expected total cost (1). The bounds hold for all prior beliefs and are nonasymptotic and complementary for various values of L and M . In Section 5 these bounds are collectively used to generalize the (information theoretic) notions of achievable communication rate [18] and error exponent [23] to the context of active sequential hypothesis testing.
- The first and second lower bounds identify the dominating terms in L and hence are useful in establishing asymptotic optimality of order-1 (due to Chernoff [17])

⁴In [17], the objective was to minimize $c\mathbb{E}[\tau] + P_e$ and the proposed policy was shown to be asymptotically optimal as $c \rightarrow 0$. It is straightforward to show that for $L = \frac{1}{c}$, this problem coincides with Problem (P) defined in this paper. However, we have chosen $\mathbb{E}[\tau] + L P_e$ as an objective function for Problem (P) because of its interpretation as the Lagrangian relaxation of an information acquisition problem in which the objective is to minimize $\mathbb{E}[\tau]$ subject to $P_e \leq \varepsilon$, where $\varepsilon > 0$ denotes the desired probability of error.

⁵In [13], Burnashev attempted to tackle the problem of active sequential hypothesis testing by Chernoff [17]. However, the sensing actions in [13] were allowed to be functions of the true hypothesis, θ , which, in general, is not observable in the active testing setting [17]. In this sense, [13] only extends Burnashev's earlier work [12] on variable-length coding over a discrete memoryless channel (DMC) with feedback to allow for more general channels.

and order-2 in L . Furthermore, from an information theoretic viewpoint, these bounds are used to characterize an upper bound on the reliability function (error exponent) at zero rate.

- The third lower bound characterizes the dominating terms of growth in the optimal expected total cost in terms of L and M simultaneously. We use this as a converse (in a fashion somewhat similar to Shannon's channel coding converse [18]) to derive an upper bound \bar{I}_{\max} on the maximum achievable information acquisition rate. Additionally, this lower bound allows us to provide an upper bound on the reliability function (error exponent) for all rates $R \in [0, \bar{I}_{\max}]$, and establish order optimality in M as a necessary condition for any policy which achieves nonzero information acquisition rate.

In addition to a lower bound on an expected number of samples, Burnashev proposed a coding scheme with two phases of operation whose performance provides a tight upper bound (in both number of messages and error probability). It is interesting to note that the scheme of Chernoff, if specialized to channel coding with feedback, coincides with the second phase of Burnashev's scheme and is of a repetition code nature. This means that while the first phase of Burnashev's scheme can achieve any information rate up to the capacity of the channel, Chernoff's one-phase scheme has a rate of information acquisition equal to zero. Inspired by Burnashev's coding scheme, we also obtain two heuristic two-phase policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$ whose nonasymptotic analysis in Proposition 2 and Theorem 3 provides two upper bounds on the optimal performance:

- Policy $\tilde{\pi}_1$ is a simple two-phase modification of Chernoff's scheme in which testing for the maximum likely hypothesis is delayed and contingent on obtaining a certain level of confidence. More specifically, in its first phase, $\tilde{\pi}_1$ selects actions in a way that all pairs of hypotheses can be distinguished from each other, while its second phase coincides with Chernoff's scheme [17] where only the pairs including the most likely hypothesis are considered. The second phase of $\tilde{\pi}_1$ ensures its asymptotic optimality in L , while its first phase in a very natural manner weakens the technical assumption in [17] in which all actions are assumed to discriminate between all hypotheses pairs or the need for the infinitely often reliance on suboptimal randomized action deployed in [17, 47].
- Policy $\tilde{\pi}_2$ is only shown to be asymptotically optimal in L under a stronger condition, which is later shown to be satisfied in the important cases of binary hypothesis testing and noisy dynamic search in Section 6, however, with the advantage that here for a fixed M the asymptotic optimality [17] can be strengthened to a higher order. In particular, in Section 5.1, we show that when $\tilde{\pi}_2$ is asymptotically optimal it achieves a bounded difference with the optimal performance. Furthermore, under a technical condition, policy $\tilde{\pi}_2$ can ensure that information acquisition occurs at a nonzero rate. Mathematically, this means that, under policy $\tilde{\pi}_2$, the expected total cost (1) grows in L and M in an order optimal fashion establishing a lower bound on the maximum achievable

information acquisition rate $\underline{L}_2 \leq \overline{I}_{\max}$ as well as a lower bound on the optimal reliability function (optimal error exponent) for all rates $R \in [0, \underline{L}_2]$.

To illustrate contributions of our work as well as highlight the rate–reliability trade-off, we treat the problem of noisy dynamic search in Section 6.2. This problem is of independent and extensive interest, and arises in a variety of fields from fault detection to whereabouts search to noisy group testing. We specialize the results obtained in the earlier sections for the general active hypothesis testing, and discuss our findings in the context of other solutions in the literature. Particularly, in the case of size-independent Bernoulli noise, the upper bound corresponding to policy $\tilde{\pi}_2$ is shown to be asymptotically tight in both L and M , hence ensuring the maximum acquisition rate and reliability simultaneously, but there is no guarantee on the tightness of the bounds for general noise models. The potentially growing gap between the lower and upper bounds obtained here, in particular, underline the significant complications of acquiring information in the general active hypothesis testing over that of (variable-length coding with feedback) [13]. For instance, while in the channel coding context the maximum information rate and reliability are fully known and match that of channel capacity and error exponent, they remain largely uncharacterized, beyond our bounds here, even in the practically relevant problem of a noisy dynamic search.

As briefly discussed in the [Introduction](#), the above results have all been obtained under an important technical assumption which is stronger than those commonly made in the literature. However, we will show that this assumption can be significantly weakened. More precisely, we show that our original technical assumption can be replaced with one that is weaker, to the best of our knowledge, than all other assumptions in the literature [1, 8, 11, 12, 32, 39, 47], and, in particular, subsumes that of [17], to obtain a set of (nonasymptotic) bounds which are looser than those obtained in Section 4. On the other hand, these looser (nonasymptotic) bounds are shown to have similar dominating terms to those obtained in Section 4, and hence ensure the validity of our asymptotic results in Section 5.

3. Dynamic programming and characterization of an optimal policy. In this section we first derive the corresponding dynamic programming (DP) equation for Problem (P). From the DP solution, we characterize an optimal policy for Problem (P).

The problem of active M -ary hypothesis testing is a partially observable Markov decision problem (POMDP) where the state is static and observations are noisy. It is known that any POMDP is equivalent to an MDP with a compact yet uncountable state space, for which the belief of the decision maker about the underlying state becomes an information state [33]. In our setup, thus, the information state at time t is the belief vector $\rho(t)$ whose i th element is the conditional probability of hypothesis H_i to be true given the initial belief and all the observations and actions up to time t , that is, $\rho_i(t) := P(\{\theta = i\} | A_0^{t-1}, Z_0^{t-1})$. Accordingly, the information

state space is defined as $\mathbb{P}(\Omega_M) := \{\boldsymbol{\rho} \in [0, 1]^M : \sum_{i=1}^M \rho_i = 1\}$ and the optimal expected total cost can be defined as follows.

DEFINITION 1. For all $\boldsymbol{\rho} \in \mathbb{P}(\Omega_M)$, let functional $V^*(\boldsymbol{\rho})$, hereafter referred to as the *optimal value function*, denote the optimal expected total cost (1) of Problem (P) given the Bayesian prior $\boldsymbol{\rho}$. In other words, $V^*(\boldsymbol{\rho}) := \min\{\mathbb{E}[\tau] + L \text{Pe}\}$ given the initial belief $\boldsymbol{\rho}$, where the minimization is taken over the stopping time τ , the sequence of actions and observations, and the declaration rule.

A general approach to solving Problem (P) is to provide a functional characterization of V^* : given V^* in its functional form, the optimal expected total cost for Problem (P) can be obtained by a simple evaluation of V^* at the initial belief $\boldsymbol{\rho}(0)$. Next we state a dynamic programming equation which characterizes V^* .

To obtain the dynamic programming equation, consider a single step of the problem. In one sensing step, the evolution of the belief vector follows Bayes' rule and is given by Φ^a , a measurable function from $\mathbb{P}(\Omega_M) \times \mathcal{Z}$ to $\mathbb{P}(\Omega_M)$ for all $a \in \mathcal{A}_M$:

$$(2) \quad \Phi^a(\boldsymbol{\rho}, z) := \left[\rho_1 \frac{q_1^a(z)}{q_{\boldsymbol{\rho}}^a(z)}, \rho_2 \frac{q_2^a(z)}{q_{\boldsymbol{\rho}}^a(z)}, \dots, \rho_M \frac{q_M^a(z)}{q_{\boldsymbol{\rho}}^a(z)} \right],$$

where $q_{\boldsymbol{\rho}}^a(z) = \sum_{i=1}^M \rho_i q_i^a(z)$, and $\Phi^a(\boldsymbol{\rho}, z) = \boldsymbol{\rho}$ if $q_{\boldsymbol{\rho}}^a(z) = 0$. In other words, if $\boldsymbol{\rho} \in \mathbb{P}(\Omega_M)$ is an a priori distribution, $\Phi^a(\boldsymbol{\rho}, z)$ gives us the posteriori distribution when sensing action a has been taken and z has been observed.

We define a Markov operator \mathbb{T}^a , $a \in \mathcal{A}_M$, such that for any measurable function $g : \mathbb{P}(\Omega_M) \rightarrow \mathbb{R}$,

$$(3) \quad (\mathbb{T}^a g)(\boldsymbol{\rho}) := \int g(\Phi^a(\boldsymbol{\rho}, z)) q_{\boldsymbol{\rho}}^a(z) dz.$$

Note that at any given information state $\boldsymbol{\rho}$, taking sensing action $a \in \mathcal{A}_M$ followed by the optimal policy results in expected total cost $1 + (\mathbb{T}^a V^*)(\boldsymbol{\rho})$, where 1 denotes the one unit of time spent to take the sensing action and collect the corresponding observation sample, and $(\mathbb{T}^a V^*)(\boldsymbol{\rho})$ is the expected value of V^* on the space of posterior beliefs; while declaration j results in expected cost $(1 - \rho_j)L$ where $(1 - \rho_j)$ is the probability that hypothesis H_j is not true, and L is the penalty of making a wrong declaration. This intuition, while relying on the compactness of $\mathbb{P}(\Omega_M)$ to treat various measurability issues, can be formalized in the following dynamic programming equation.

FACT 1 (Proposition 9.8 in [7]). The optimal value function V^* satisfies the following fixed point equation:

$$(4) \quad V^*(\boldsymbol{\rho}) = \min \left\{ 1 + \min_{a \in \mathcal{A}_M} (\mathbb{T}^a V^*)(\boldsymbol{\rho}), \min_{j \in \Omega_M} (1 - \rho_j)L \right\}.$$

DEFINITION 2. A Markov stationary *policy* is a stochastic kernel from the information state space $\mathbb{P}(\Omega_M)$ to $\mathcal{A}_M \cup \{d\}$ describing the conditional distribution on sensing actions $A(t)$, $t = 0, 1, \dots, \tau - 1$ and stopping time τ (the choice of declaration d marks the stopping time τ). In other words, under policy π , the probability that action a is selected at belief state ρ is given by $\pi(a|\rho)$.

As shown in Corollary 9.12.1 in [7], equation (4) provides a characterization of an optimal Markov stationary deterministic policy π^* for Problem (P) as follows: sensing action $a^* = \arg \min_{a \in \mathcal{A}_M} (\mathbb{T}^a V^*)(\rho)$ is the least costly sensing action, resulting in $1 + \min_{a \in \mathcal{A}_M} (\mathbb{T}^a V^*)(\rho)$, hence is the optimal action to take unless wrongly declaring H_{i^*} , where $i^* = \arg \min_{j \in \Omega_M} (1 - \rho_j)L$, is even less costly, in which case it is optimal to retire and declare H_{i^*} as the true hypothesis.

REMARK 1. It follows from (4) that if $\min_{j \in \Omega_M} (1 - \rho_j)L \leq 1$, then we have a full characterization of $V^*(\rho) = \min_{j \in \Omega_M} (1 - \rho_j)L$ and the optimal policy. Therefore, the region of interest in our analysis is restricted to $L > 1$ and $\mathbb{P}_L(\Omega_M) := \{\rho \in \mathbb{P}(\Omega_M) : \min_{j \in \Omega_M} (1 - \rho_j)L > 1\}$.

Before we close this section, we provide the following lemma.

LEMMA 1. Suppose there exist $\beta > 0$ and a functional $V : \mathbb{P}(\Omega_M) \rightarrow \mathbb{R}_+$ such that for all belief vectors $\rho \in \mathbb{P}(\Omega_M)$,

$$V(\rho) \leq \min \left\{ \beta + \min_{a \in \mathcal{A}_M} (\mathbb{T}^a V)(\rho), \min_{j \in \Omega_M} (1 - \rho_j)\beta L \right\}.$$

Then $V^*(\rho) \geq \frac{1}{\beta} V(\rho)$ for all $\rho \in \mathbb{P}(\Omega_M)$.

The proof is provided in the supplemental article [44], Section 1.

4. Performance bounds. In lieu of numerical approximation of or derivation of a closed form for V^* , in Section 4.1 we use Lemma 1 to find lower bounds for the value function V^* . In Section 4.2 we analyze two heuristic schemes to achieve upper bounds for V^* .

We have the following technical assumptions:

ASSUMPTION 1. For any two hypotheses $i, j \in \Omega_M$, $i \neq j$, there exists an action a , $a \in \mathcal{A}_M$, such that $D(q_i^a \| q_j^a) > 0$.

ASSUMPTION 2. There exists $\xi_M < \infty$ such that

$$\max_{i, j \in \Omega_M} \max_{a \in \mathcal{A}_M} \sup_{z \in \mathcal{Z}} \log \frac{q_i^a(z)}{q_j^a(z)} \leq \xi_M.$$

Assumption 1 ensures the possibility of discrimination between any two hypotheses, hence ensuring Problem (P) has a meaningful solution. Assumption 2

implies that no two hypotheses are fully distinguishable using a single observation sample. Assumption 2 is a technical one which enables our nonasymptotic characterizations, however, in Section 7 we discuss the consequence of weakening this assumption in detail.

4.1. Lower bounds for V^* .

THEOREM 1. Under Assumption 1 and for $L > 1$ and $\rho \in \mathbb{P}_L(\Omega_M)$,

$$V^*(\rho) \geq \underline{V}_1(\rho) := \left[\sum_{i=1}^M \rho_i \max_{j \neq i} \frac{\log((1 - L^{-1})/L^{-1}) - \log(\rho_i/\rho_j)}{\max_{a \in \mathcal{A}_M} D(q_i^a \| q_j^a)} - K'_1 \right]^+,$$

where K'_1 is a constant independent of L whose closed form is given in the supplemental article [44], equation (144).

The proof of Theorem 1 is provided in Appendix.

Following Chernoff’s approach (Theorem 2 in [17]), and for large values of L , the lower bound can be tightened as follows:

PROPOSITION 1. Under Assumptions 1 and 2, and for $L > 1$, $\rho \in \mathbb{P}_L(\Omega_M)$, and arbitrary $\delta \in (0, 1)$,

$$V^*(\rho) \geq \left[\sum_{i=1}^M \rho_i \frac{[(1 - \delta) \log(L/(K' \log 2L)) - \max_{j \neq i} \log(\rho_i/\rho_j)]^+}{\max_{\lambda \in \mathbb{P}(\mathcal{A}_M)} \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \| q_j^a) + \delta} \times \left(1 - \frac{2M(K' \log 2L/L)^\delta}{\rho_i} \right) - \frac{M\xi_M^2}{\delta^2} \right]^+,$$

where K' is a constant independent of δ and L whose closed form is given in the supplemental article [44], equation (81).

The proof of Proposition 1 is provided in the supplemental article [44], Section 5.1.

Next we provide another lower bound which is more appropriate for large values of M . Let $I(\rho; q_\rho^a) = H(\rho) - (\mathbb{T}^a H)(\rho)$ denote the mutual information between $\theta \sim \rho$ and observation Z under action a . Let $D_{\max}(M) := \max_{i,j \in \Omega_M} \max_{a \in \mathcal{A}_M} D(q_i^a \| q_j^a)$, $I_{\max}(M) := \max_{a \in \mathcal{A}_M} \max_{\hat{\rho} \in \mathbb{P}(\Omega_M)} I(\hat{\rho}; q_{\hat{\rho}}^a)$, and $\alpha(L, M) := \frac{M-1}{M-1+2^L I_{\max}(M)}$.

THEOREM 2. Under Assumption 1 and for $L > 1$ and $\rho \in \mathbb{P}_L(\Omega_M)$,

$$V^*(\rho) \geq \left[\frac{H(\rho) - H([\alpha(L, M), 1 - \alpha(L, M)]) - \alpha(L, M) \log(M - 1)}{I_{\max}(M)} + \alpha(L, M)L \right]^+.$$

Furthermore, under Assumptions 1 and 2, and for $L > \max\{1, \frac{\log M}{I_{\max}(M)}\}$ and arbitrary $\delta \in (0, 0.5]$,

$$\begin{aligned} V^*(\boldsymbol{\rho}) &\geq \underline{V}_2(\boldsymbol{\rho}) \\ &:= \left[\frac{H(\boldsymbol{\rho}) - H([\delta, 1 - \delta]) - \delta \log(M - 1)}{I_{\max}(M)} \right. \\ &\quad \left. + \frac{\log((1 - L^{-1})/L^{-1}) - \log((1 - \delta)/\delta) - \xi_M}{D_{\max}(M)} \right]^+ \\ &\quad \times \mathbf{1}_{\{\max_{i \in \Omega_M} \rho_i \leq 1 - \delta\}} - K'_2 \end{aligned}$$

where K'_2 is a constant independent of δ and L whose closed form is given in the supplemental article [44], equation (151).⁶

The proof of Theorem 2 is provided in Appendix.

Theorem 2 can be used to show that when $L < \frac{\log M}{I_{\max}(M)}$, Problem (P) will have a trivial solution. The precise statement is given by the following corollary.

COROLLARY 1. *Let $L < \frac{\log M}{I_{\max}(M)}$, and suppose the decision maker has a uniform prior belief about the hypotheses. For sufficiently large M , the optimal policy randomly guesses the true hypothesis without collecting any observation, hence, P_e , the probability of making a wrong declaration, approaches $1 - \frac{1}{M}$.*

The proof of Corollary 1 is provided in the supplemental article [44], Section 2.1.

REMARK 2. The lower bounds in Theorems 1 and 2 can be explained by the following intuition: for any measure of uncertainty $U: \mathbb{P}(\Omega_M) \rightarrow \mathbb{R}_+$, the number of samples required to reduce the uncertainty down to a target level U_{target} has to be at least $\frac{U(\boldsymbol{\rho}(0)) - U_{\text{target}}}{\Delta_{\max}(U)}$, where $\Delta_{\max}(U)$ is the maximum amount of reduction in U associated with a single sample, that is, $\Delta_{\max}(U) = \max_{a \in \mathcal{A}_M} \max_{\boldsymbol{\rho} \in \mathbb{P}(\Omega_M)} \{U(\boldsymbol{\rho}) - (\mathbb{T}^a U)(\boldsymbol{\rho})\}$. The lower bound in Theorem 1 is associated with such a lower bound when taking U to be the log-likelihood function, while the lower bound in Theorem 2 is associated with setting U to be the Shannon entropy.

⁶As it will be discussed in Section 5.2, K'_2 can be selected independent of M as well if $\sup_M \xi_M < \infty$.

4.2. *Upper bounds for V^* .* Next we propose two Markov policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$. Policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$ have two operational phases. Phase 1 is the phase in which the belief about all hypotheses is below a certain threshold, while in phase 2, the belief about one of the hypotheses has passed that threshold and actions are selected in favor of that particular hypothesis. The difference between the two policies is in the actions they take in each phase.

First we describe policy $\tilde{\pi}_1$. Let μ_0 and $\mu_i, i \in \Omega_M$, be vectors in $\mathbb{P}(\mathcal{A}_M)$ such that

$$\begin{aligned} \mu_0 &:= \arg \max_{\lambda \in \mathbb{P}(\mathcal{A}_M)} \min_{i \in \Omega_M} \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \| q_j^a), \\ \mu_i &:= \arg \max_{\lambda \in \mathbb{P}(\mathcal{A}_M)} \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \| q_j^a) \quad \forall i \in \Omega_M. \end{aligned}$$

Moreover, let μ_{0a} and μ_{ia} denote elements of μ_0 and μ_i corresponding to $a \in \mathcal{A}_M$, respectively. Consider a threshold $\tilde{\rho}, \tilde{\rho} > \frac{1}{2}$. Markov (randomized) policy $\tilde{\pi}_1$ is defined as follows:⁷

- If $\rho_i \geq 1 - L^{-1}$, retire and select H_i as the true hypothesis.
- If $\rho_i \in [\tilde{\rho}, 1 - L^{-1})$, then
 - $\tilde{\pi}_1(a|\rho) = \mu_{ia} \forall a \in \mathcal{A}_M$.
- If $\rho_i < \min\{\tilde{\rho}, 1 - L^{-1}\}$ for all $i \in \Omega_M$, then
 - $\tilde{\pi}_1(a|\rho) = \mu_{0a} \forall a \in \mathcal{A}_M$.

In [17], Chernoff proposed a policy that, at each time t , selects action a with probability μ_{i^*a} , where $i^* = \arg \max_{i \in \Omega_M} \rho_i(t)$ denotes the most likely true hypothesis. In other words, $\tilde{\pi}_1$ coincides with Chernoff’s scheme in its second phase and ensures its asymptotic optimality in L , while its first phase in a very natural manner relaxes the technical assumption in [17] where all actions were required to discriminate between all hypotheses pairs. Following Chernoff’s approach (Theorem 1 in [17]), we can analyze the performance of policy $\tilde{\pi}_1$ and obtain the following upper bound for V^* .

For notational simplicity, let

$$\begin{aligned} I_{\mu_0}(M) &:= \min_{i \in \Omega_M} \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \mu_{0a} D(q_i^a \| q_j^a), \\ I_1(M) &:= \left(\frac{\log M + 4\xi_M}{\min_{i \in \Omega_M} \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \mu_{ja} D(q_i^a \| q_j^a)} \right)^{-2} I_{\mu_0}(M), \\ D_{\mu_i}(M) &:= \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \mu_{ia} D(q_i^a \| q_j^a) \quad \forall i \in \Omega_M. \end{aligned}$$

⁷Policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$ are not unique; they each represent a class of parameterized policies. In fact, the tilde in $\tilde{\pi}_1$ and $\tilde{\pi}_2$ has been chosen to emphasize the dependency of these policies on the threshold/parameter $\tilde{\rho}$.

PROPOSITION 2. Under Assumptions 1 and 2, and for $L > 1$, $\rho \in \mathbb{P}_L(\Omega_M)$, and arbitrary $\iota \in (0, 1)$,

$$\begin{aligned} V^*(\rho) &\leq \bar{V}_1(\rho) \\ &:= \frac{H(\rho) + \log M + \log(\tilde{\rho}/(1 - \tilde{\rho}))}{I_1(M)}(1 + \iota) + \sum_{i=1}^M \rho_i \frac{\log L}{D_{\mu_i}(M)}(1 + \iota) \\ &\quad + M \left(2 + \frac{1}{((\iota/2)/(1 + \iota))^5 (I_1(M)/(2\xi_M))^4} \right) \\ &\quad \times \left(L \left(1 - \max_{j \in \Omega_M} \rho_j \right) \right)^{-((\iota^3/(1 + \iota)^2) I_1^2(M)/(4\xi_M^3))} + 2. \end{aligned}$$

The proof is based on a performance analysis of policy $\tilde{\pi}_1$ and is provided in the supplemental article [44], Section 5.2.

Next we describe policy $\tilde{\pi}_2$. Let η_0 and η_i , $i \in \Omega_M$, be vectors in $\mathbb{P}(\mathcal{A}_M)$ such that

$$\begin{aligned} \eta_0 &:= \arg \max_{\lambda \in \mathbb{P}(\mathcal{A}_M)} \min_{i \in \Omega_M} \min_{\hat{\rho} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a D \left(q_i^a \left\| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a \right. \right), \\ \eta_i &:= \arg \max_{\lambda \in \mathbb{P}(\mathcal{A}_M)} \min_{\hat{\rho} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a D \left(q_i^a \left\| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a \right. \right) \quad \forall i \in \Omega_M. \end{aligned}$$

Moreover, let η_{0a} and η_{ia} denote elements of η_0 and η_i corresponding to $a \in \mathcal{A}_M$, respectively. Consider a threshold $\tilde{\rho}$, $\tilde{\rho} > \frac{1}{2}$. Markov (randomized) policy $\tilde{\pi}_2$ is defined as follows:

- If $\rho_i \geq 1 - L^{-1}$, retire and select H_i as the true hypothesis.
- If $\rho_i \in [\tilde{\rho}, 1 - L^{-1})$, then
 - $\tilde{\pi}_2(a|\rho) = \eta_{ia} \quad \forall a \in \mathcal{A}_M$.
- If $\rho_i < \min\{\tilde{\rho}, 1 - L^{-1}\}$ for all $i \in \Omega_M$, then
 - $\tilde{\pi}_2(a|\rho) = \eta_{0a} \quad \forall a \in \mathcal{A}_M$.

For notational simplicity, let

$$\begin{aligned} I_{\eta_0}(M) &:= \min_{i \in \Omega_M} \min_{\hat{\rho} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \eta_{0a} D \left(q_i^a \left\| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a \right. \right), \\ I_{\eta, \tilde{\rho}}(M) &:= \min_{i \in \Omega_M} \min_{k \neq i} \min_{\hat{\rho} : \hat{\rho}_k \geq \tilde{\rho}} \sum_{a \in \mathcal{A}_M} \eta_{ka} D \left(q_i^a \left\| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a \right. \right), \\ I_2(M) &:= \min\{I_{\eta_0}(M), I_{\eta, \tilde{\rho}}(M)\}, \\ D_{\eta_i}(M) &:= \min_{\hat{\rho} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \eta_{ia} D \left(q_i^a \left\| \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a \right. \right) \quad \forall i \in \Omega_M. \end{aligned}$$

THEOREM 3. *Under Assumptions 1 and 2, and for $L > 1$ and any $\rho \in \mathbb{P}_L(\Omega_M)$,*

$$V^*(\rho) \leq \bar{V}_2(\rho) := \frac{H(\rho) + \log(\tilde{\rho}/(1 - \tilde{\rho})) + \xi_M + \log e}{I_2(M)} + \sum_{i=1}^M \rho_i \frac{\log L}{D_{\eta_i}(M)} + 1.$$

The proof is based on a performance analysis of policy $\tilde{\pi}_2$ and is provided in the [Appendix](#).

5. Asymptotic analysis and consequences. In this section we state and discuss the consequence of the bounds obtained in Section 4 in asymptotically large L and M . Note that Table 1 provides a list of the notation introduced in Section 4.

5.1. *Order and asymptotic optimality in L .* The lower and upper bounds provided in Section 4 can be applied to establish the order optimality and asymptotic optimality of the proposed policies as defined below. Let $V_\pi(\rho)$ denote the value

TABLE 1
Summary of notation

Notation	Description
$I_{\max}(M)$	$\max_{a \in \mathcal{A}_M} \max_{\hat{\rho} \in \mathbb{P}(\Omega_M)} I(\hat{\rho}; q_{\hat{\rho}}^a)$
$D_{\max}(M)$	$\max_{i,j \in \Omega_M} \max_{a \in \mathcal{A}_M} D(q_i^a \ q_j^a)$
μ_0	$\arg \max_{\lambda \in \mathbb{P}(\mathcal{A}_M)} \min_{i \in \Omega_M} \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \ q_j^a)$
μ_i	$\arg \max_{\lambda \in \mathbb{P}(\mathcal{A}_M)} \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \ q_j^a)$
$I_{\mu_0}(M)$	$\min_{i \in \Omega_M} \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \mu_{0a} D(q_i^a \ q_j^a)$
$I_1(M)$	$(\frac{\log M + 4\xi_M}{\min_{i \in \Omega_M} \min_{j \neq i} \sum_{a \in \mathcal{A}_M} \mu_{ja} D(q_i^a \ q_j^a)})^{-2} I_{\mu_0}(M)$
$D_{\mu_i}(M)$	$\min_{j \neq i} \sum_{a \in \mathcal{A}_M} \mu_{ia} D(q_i^a \ q_j^a)$
η_0	$\arg \max_{\lambda \in \mathbb{P}(\mathcal{A}_M)} \min_{i \in \Omega_M} \min_{\hat{\rho} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \ \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a)$
η_i	$\arg \max_{\lambda \in \mathbb{P}(\mathcal{A}_M)} \min_{\hat{\rho} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a D(q_i^a \ \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a)$
$I_{\eta_0}(M)$	$\min_{i \in \Omega_M} \min_{\hat{\rho} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \eta_{0a} D(q_i^a \ \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a)$
$I_{\eta, \tilde{\rho}}(M)$	$\min_{i \in \Omega_M} \min_{k \neq i} \min_{\hat{\rho}: \hat{\rho}_k \geq \tilde{\rho}} \sum_{a \in \mathcal{A}_M} \eta_{ka} D(q_i^a \ \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a)$
$I_2(M)$	$\min\{I_{\eta_0}(M), I_{\eta, \tilde{\rho}}(M)\}$
$D_{\eta_i}(M)$	$\min_{\hat{\rho} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \eta_{ia} D(q_i^a \ \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a)$

function for policy π , that is, the expected total cost achieved by policy π when the initial belief is ρ .

DEFINITION 3. For fixed M , policy π is referred to as *order optimal* in L if for all $\rho \in \mathbb{P}(\Omega_M)$,

$$\lim_{L \rightarrow \infty} \frac{V_\pi(\rho) - V^*(\rho)}{V_\pi(\rho)} < 1.$$

DEFINITION 4. For fixed M , policy π is referred to as *asymptotically optimal of order-1* in L if for all $\rho \in \mathbb{P}(\Omega_M)$,

$$\lim_{L \rightarrow \infty} \frac{V_\pi(\rho) - V^*(\rho)}{V_\pi(\rho)} = 0.$$

DEFINITION 5. For fixed M , policy π is referred to as *asymptotically optimal of order-2* in L if for all $\rho \in \mathbb{P}(\Omega_M)$, there exists a constant B independent of L such that

$$V_\pi(\rho) - V^*(\rho) \leq B.$$

REMARK 3. It is clear from the definitions above that order optimality is weaker than asymptotic optimality of order-1, while asymptotic optimality of order-2 is the strongest notion. The notion of asymptotic optimality of order-1 was first introduced in [17], which naturally motivates the extension of higher orders.

The next corollary establishes order and asymptotic optimality of our proposed policies.

COROLLARY 2. Under Assumptions 1 and 2, policy $\tilde{\pi}_1$ is asymptotically optimal of order-1 in L . Furthermore, policy $\tilde{\pi}_2$ attains asymptotic optimality of order-2 in L if

$$(5) \quad \min_{j \neq i} \max_{a \in \mathcal{A}_M} D(q_i^a \| q_j^a) = D_{\eta_i}(M) \quad \forall i \in \Omega_M.$$

PROOF. Using Proposition 1 and by setting $\delta = (\log L)^{-1/3}$, we obtain

$$(6) \quad V^*(\rho) \geq \sum_{i=1}^M \rho_i \frac{\log L}{D_{\mu_i}(M)} + O((\log L)^{2/3}).$$

On the other hand, from Proposition 2 and by setting $\iota = (\log L)^{-1/4}$, we get

$$(7) \quad V^*(\rho) \leq \sum_{i=1}^M \rho_i \frac{\log L}{D_{\mu_i}(M)} + O((\log L)^{3/4}).$$

The proof of the first part of the corollary simply follows from Definition 4, inequality (6) and (7).

Similarly, the proof of the second part of the corollary follows from Definition 5, Theorems 1 and 3. \square

5.2. *Order and asymptotic optimality in both L and M .* As mentioned in Section 2.2, one of the main drawbacks of Chernoff’s asymptotic optimality notion was his neglecting the complementary role of parameter M . In particular, the notion of asymptotic optimality in L falls short in showing the tension between using an (asymptotically) large number of samples to discriminate among a few hypotheses with (asymptotically) high accuracy or an (asymptotically) large number of hypotheses with a lower degree of accuracy. In this section we address this issue by analyzing the bounds when L and M are both asymptotically large. More specifically, we consider a sequence of problems indexed by parameter M in which the set of actions and observation kernels grow monotonically as M increases, that is, for all $M < M'$,

$$(8) \quad \mathcal{A}_M \subseteq \mathcal{A}_{M'} \quad \text{and} \quad \{q_i^a(\cdot)\}_{i \in \Omega_M, a \in \mathcal{A}_M} \subseteq \{q_i^a(\cdot)\}_{i \in \Omega_{M'}, a \in \mathcal{A}_{M'}}.$$

Recall the notation listed in Table 1. Also, let $D_1(M)$ and $D_2(M)$ denote, respectively, the harmonic mean of $\{D_{\mu_i}(M)\}_{i \in \Omega_M}$ and $\{D_{\eta_i}(M)\}_{i \in \Omega_M}$, that is,

$$(9) \quad D_1(M) = M \left(\sum_{i=1}^M \frac{1}{D_{\mu_i}(M)} \right)^{-1}, \quad D_2(M) = M \left(\sum_{i=1}^M \frac{1}{D_{\eta_i}(M)} \right)^{-1}.$$

Moreover, let

$$(10) \quad \bar{I}_{\max} := \sup_M I_{\max}(M), \quad \bar{D}_{\max} := \sup_M D_{\max}(M),$$

$$(11) \quad \underline{I}_{\max} := \inf_M I_{\max}(M), \quad \underline{D}_{\max} := \inf_M D_{\max}(M),$$

$$(12) \quad \underline{I}_2 := \inf_M I_2(M), \quad \underline{D}_2 := \inf_M D_2(M).$$

By the definition and from (8), $D_{\max}(M)$ and $I_{\max}(M)$ are nondecreasing in M . Furthermore, from Jensen’s inequality,

$$(13) \quad \begin{aligned} I_{\max}(M) &= \max_{a \in \mathcal{A}_M} \max_{\hat{\rho} \in \mathbb{P}(\Omega_M)} \sum_{i=1}^M \hat{\rho}_i D \left(q_i^a \parallel \sum_{j=1}^M \hat{\rho}_j q_j^a \right) \\ &\leq \max_{a \in \mathcal{A}_M} \max_{\hat{\rho} \in \mathbb{P}(\Omega_M)} \sum_{i=1}^M \hat{\rho}_i \sum_{j=1}^M \hat{\rho}_j D(q_i^a \parallel q_j^a) \\ &\leq \max_{a \in \mathcal{A}_M} \max_{i, j \in \Omega_M} D(q_i^a \parallel q_j^a) = D_{\max}(M) \end{aligned}$$

and by Assumption 2, we have⁸

$$(14) \quad D_{\max}(M) \leq \max_{i,j \in \Omega_M} \max_{a \in \mathcal{A}_M} \sup_{z \in \mathcal{Z}} \log \frac{q_i^a(z)}{q_j^a(z)} \leq \xi_M.$$

Similarly, $I_2(M) \leq D_{\eta_i}(M) \leq D_{\mu_i}(M) \leq D_{\max}(M) \leq \xi_M, \forall i \in \Omega_M$, for all M . Since $D_{\max}(M)$ and $I_{\max}(M)$ are nondecreasing in M , we have $\underline{D}_{\max} = D_{\max}(2)$, $\bar{D}_{\max} = \lim_{M \rightarrow \infty} D_{\max}(M)$, $\underline{I}_{\max} = I_{\max}(2)$ and $\bar{I}_{\max} = \lim_{M \rightarrow \infty} I_{\max}(M)$.

Furthermore, to ensure that the distance between the observation kernels remains bounded as M increases (and $\bar{D}_{\max} < \infty$), we consider the following assumption:

ASSUMPTION 3. There exists $\xi < \infty$ such that

$$\sup_M \xi_M \leq \xi.$$

This assumption allows us to specialize Theorem 2 as follows.

COROLLARY 3. Let $\rho_{u,M}$ denote a uniform prior on the set of hypotheses Ω_M . Under Assumptions 1, 2 and 3, and for $\delta = \frac{1}{\log 2ML}$ and $L > \max\{2, \frac{\log M}{\bar{I}_{\max}(M)}\}$,

$$\underline{V}_2(\rho_{u,M}) \geq \left[\frac{\log M - 2}{\bar{I}_{\max}} + \frac{\log((1 - L^{-1})/L^{-1})}{\bar{D}_{\max}} - \frac{\log \log LM + \xi}{\underline{D}_{\max}} - K'_2 \right]^+,$$

where K'_2 is a positive constant independent of L and M .

The proof of Corollary 3 is provided in the supplemental article [44], Section 2.2.

The next definition extends the notions of order and asymptotic optimality defined in Section 5.1 to the case where M increases as well.

DEFINITION 6. Policy π is referred to as *order optimal* and *asymptotically optimal of order-1* in L and M if, respectively,⁹

$$\lim_{L,M \rightarrow \infty} \frac{V_\pi(\rho_{u,M}) - V^*(\rho_{u,M})}{V_\pi(\rho_{u,M})} < 1, \quad \lim_{L,M \rightarrow \infty} \frac{V_\pi(\rho_{u,M}) - V^*(\rho_{u,M})}{V_\pi(\rho_{u,M})} = 0.$$

⁸Inequality (14) holds true even if Assumption 2 is replaced by a more general assumption such as those suggested in Section 7.

⁹Note that unlike Definitions 3–5 where we considered the performance gap between policy π and the optimal policy π^* for all values of $\rho \in \mathbb{P}(\Omega_M)$, here we consider the performance gap specifically at the uniform vector in the information state space.

COROLLARY 4. Under Assumptions 1, 2 and 3, for $L > \frac{\log M}{I_{\max}(M)}$, and if $\underline{I}_2 > 0$, policy $\tilde{\pi}_2$ is order optimal in L and M . Furthermore, if $\underline{I}_{\max} = \underline{I}_2$ and $\underline{D}_{\max} = \underline{D}_2$, policy $\tilde{\pi}_2$ is asymptotically optimal of order-1 in L and M .

PROOF. The proof follows from Definition 6, Corollary 3 and Theorem 3. \square

5.3. *Information acquisition rate and reliability.* In this section we explain the primal (constrained) version of Problem (P), referred to as Problem (P'), and use the obtained bounds in Section 4 to extend the (information theoretic) notions of achievable communication rate and error exponent to the context of active sequential hypothesis testing.

PROBLEM (P') (Information acquisition problem). Consider a sequence of active hypothesis testing problems indexed by parameter M (i.e., the number of hypotheses of interest), action space \mathcal{A}_M and observation kernels $\{q_i^a(\cdot)\}_{i \in \Omega_M, a \in \mathcal{A}_M}$: a Bayesian decision maker with uniform prior belief $\rho(0) = \rho_{u,M}$ is responsible to find the true hypothesis with the objective to

$$(15) \quad \text{minimize } \mathbb{E}[\tau] \text{ subject to } \text{Pe} \leq \varepsilon,$$

where τ is the stopping time at which the decision maker retires, Pe is the probability of making a wrong declaration, and $\varepsilon > 0$ denotes the desired probability of error. Furthermore, let the set of actions and observation kernels grow monotonically as M increases, that is, for all $M < M'$,

$$(16) \quad \mathcal{A}_M \subseteq \mathcal{A}_{M'} \quad \text{and} \quad \{q_i^a(\cdot)\}_{i \in \Omega_M, a \in \mathcal{A}_M} \subseteq \{q_i^a(\cdot)\}_{i \in \Omega_{M'}, a \in \mathcal{A}_{M'}}.$$

Let $\mathbb{E}_\pi[\tau]$ and Pe_π denote, respectively, the expected stopping time (or, equivalently, the expected number of collected samples) and the probability of error under policy π . Following the notation in [49], we define $M_\pi(t, \varepsilon)$ as the maximum number of hypotheses among which policy π can find the true hypothesis with $\mathbb{E}_\pi[\tau] \leq t$ and $\text{Pe}_\pi \leq \varepsilon$. Policy π is said to achieve information acquisition rate $R > 0$ with reliability (also known as error exponent) $E > 0$ if

$$(17) \quad \lim_{t \rightarrow \infty} \frac{1}{t} \log M_\pi(t, 2^{-Et}) = R.$$

For a fixed number of hypotheses M , hence at information acquisition rate $R = 0$, policy π is said to achieve reliability $E > 0$ if

$$(18) \quad \lim_{t \rightarrow \infty} \frac{-1}{t} \log \text{Pe}_\pi(t, M) = E,$$

where $\text{Pe}_\pi(t, M)$ is the minimum probability of error that policy π can guarantee for M hypotheses with the constraint $\mathbb{E}_\pi[\tau] \leq t$.

The reliability function $E(R)$ is defined as the maximum achievable error exponent at information acquisition rate R .

Before we proceed with the upper and lower bounds on the maximum achievable information acquisition rate and the optimal reliability function, we refer the reader to Table 1 for the list of notation introduced in Section 4. Also recall that $D_1(M)$ and $D_2(M)$ denote, respectively, the harmonic mean of $\{D_{\mu_i}(M)\}_{i \in \Omega_M}$ and $\{D_{\eta_i}(M)\}_{i \in \Omega_M}$.

COROLLARY 5. *For any given fixed M (rate $R = 0$), no policy can achieve reliability higher than $D_1(M)$. Also, no policy can achieve positive reliability $E > 0$ at rates higher than \bar{I}_{\max} . Furthermore,*

$$(19) \quad E(R) \leq \bar{D}_{\max} \left(1 - \frac{R}{\bar{I}_{\max}} \right), \quad R \in (0, \bar{I}_{\max}).$$

REMARK 4. Corollary 5 establishes an upper bound, \bar{I}_{\max} , on the maximum achievable information acquisition rate. As shown in the supplemental article [44], Section 3.3, this result can be strengthened to show that no policy can achieve diminishing error probability at rates higher than \bar{I}_{\max} .

COROLLARY 6. *For fixed M , hence at rate $R = 0$, a policy π can achieve the maximum reliability, that is, $E = D_1(M)$, if and only if it is asymptotically optimal (of order-1 or higher) in L . Furthermore, a policy π can achieve a nonzero rate $R > 0$ with nonzero reliability $E > 0$ only if it is order optimal in L and M .*

Corollary 6 implies that for fixed M , hence at $R = 0$, policies $\tilde{\pi}_1$ and π^* achieve the optimal error exponent, while policy $\tilde{\pi}_2$ might or might not [depending on condition (5)]. Furthermore, Corollary 6, in effect, underlines the deficiency of characterizing the solution to Problems (P) in terms of L in isolation from M , hence, Chernoff's notion of asymptotic optimality (solely in L). In particular, an order optimal policy can achieve nonzero rate and reliability simultaneously, an improvement over $\tilde{\pi}_1$ (and all extensions of [17]).

COROLLARY 7. *Policy $\tilde{\pi}_2$ achieves rate $R \in [0, \underline{I}_2]$ with reliability E if*

$$(20) \quad E \leq \underline{D}_2 \left(1 - \frac{R}{\underline{I}_2} \right).$$

Figure 1 summarizes the results above. The upper bound on the reliability function is shown in red. Policy $\tilde{\pi}_1$ achieves the optimal reliability $D_1(M)$ for fixed M (at $R = 0$) with no provable guarantee for $R > 0$ (this point is shown in green), while policy $\tilde{\pi}_2$ ensures an exponentially decaying error probability (the error exponent is shown in blue) for $R \in [0, \underline{I}_2]$.

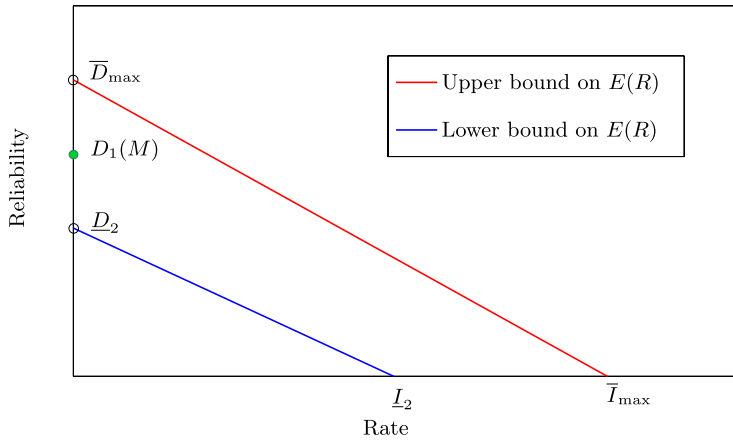


FIG. 1. Lower and upper bounds on the optimal reliability function $E(R)$.

REMARK 5. It can be shown that any optimal policy π^* for Problem (P) also achieves any rate $R \in [0, \underline{I}_2]$ with reliability E satisfying (20) for Problem (P').

The proofs of all the results in this section are provided in the supplemental article [44], Section 3, and are based on the fact that Problem (P) can be viewed as a Lagrangian relaxation of Problem (P'). It is somewhat intuitive that as $L \rightarrow \infty$ the solution of Problem (P) is closely related to that of Problem (P') when $\varepsilon \rightarrow 0$. The following lemma makes this intuition precise.

LEMMA 2. Let $\mathbb{E}[\tau_\varepsilon^*]$ denote the minimum expected number of samples required to achieve $\text{Pe} \leq \varepsilon$. We have

$$(21) \quad \mathbb{E}[\tau_\varepsilon^*] \geq (1 - \varepsilon L)(V^*(\boldsymbol{\rho}(0)) - 1),$$

where $V^*(\boldsymbol{\rho}(0))$ is the optimal solution to Problem (P) for prior belief $\boldsymbol{\rho}(0)$ and penalty of wrong declaration L .

Given the above connection, Corollary 5 follows readily from the lower bounds obtained in Proposition 1 and Theorem 2 (in particular, its Corollary 3), and Corollaries 6 and 7 follow from the upper bounds given by Proposition 2 and Theorem 3.

6. Examples. In this section we consider important special cases of the active hypothesis testing to provide some intuition about the conditions of Corollaries 2 and 4, and, in particular, establish the order-2 asymptotic optimality of $\tilde{\pi}_2$ for a fixed value of M and rate–reliability optimality of policy $\tilde{\pi}_2$.

6.1. *Binary hypothesis testing.* Consider Problem (P) for $M = 2$. In this setting, policies $\tilde{\pi}_1$ and $\tilde{\pi}_2$ are equivalent and by Corollary 2, both policies are asymptotically optimal of order-1 in L . Asymptotic optimality of order-2 of $\tilde{\pi}_1$ and $\tilde{\pi}_2$ is also verified from Corollary 2 since equality (5) holds trivially for $M = 2$. Furthermore, we obtain

$$V^*(\rho) = \rho_1 \frac{\log L - \log(\rho_1/\rho_2)}{\max_{a \in \mathcal{A}_M} D(q_1^a \| q_2^a)} + \rho_2 \frac{\log L - \log(\rho_2/\rho_1)}{\max_{a \in \mathcal{A}_M} D(q_2^a \| q_1^a)} + O(1).$$

The problem of reliability (error exponent) associated with passive binary hypothesis testing with fixed-length (nonsequential) as well as variable-length (sequential) sample size has been studied by [10, 19, 25]. The generalization to channel coding with feedback with two messages was addressed in [4, 5, 46]. Recently, the authors in [26] and [50] have generalized this problem for fixed-length and variable-length sample size, respectively, to the active binary hypothesis testing in the non-Bayesian context, and identified the error exponent corresponding to both error types. Our work provides nonasymptotic bounds as well as an asymptotic optimal solution in a total cost and Bayesian sense, and is consistent with the findings in [50].

6.2. *Noisy dynamic search.* Consider the problem of sequentially searching for a single target in M locations where the goal is to find the target quickly and accurately. In each step, the player can inspect an allowable combination of the locations, and the outcome of the inspection is noisy. This problem is closely related to the problems of *fault detection*, *whereabouts search* and *group testing*. In fault detection, the objective is to determine the faulty component in a system known to have one failed component [15, 42]. In whereabouts search, the goal is to find an object which is hidden in one of M boxes, where it is usually assumed that there is no *false alarm*, that is, the outcome of inspecting box i is always 0 if no object is present, and is a Bernoulli random variable with a known parameter otherwise [30, 56]. In group testing, the goal is to locate the nonzero element¹⁰ of a vector in \mathbb{R}^M with a possible noisy linear measurement of the vector [16, 52]. One possible search strategy for these problems is the maximum likelihood policy. In the case of fault detection/whereabouts search, this policy is equivalent to one that inspects a segment with the highest probability of having the faulty component/hidden object, while in the case of group testing, it is equivalent to measuring the most likely nonzero element of the vector. However, as the number of segments or the dimension of vectors, M , increases, the scheme becomes impractical. In such a case, it is more intuitive to initially follow a noisy binary search [14, 29, 48] and narrow down the search to single segments only after we have collected sufficient information supporting the presence of the target in those segments [51, 55].

¹⁰Group testing with $d > 1$ nonzero elements is also a special case of active hypothesis testing with $\binom{M}{d}$ hypotheses (possible configurations).

In this section, we first consider the problem of a noisy dynamic search with size-dependent Bernoulli noise whose special cases have been independently studied in [14, 16, 29, 30, 42, 48, 52, 56].¹¹ Remark 6 at the end of this section discusses a generalization for the symmetric noise model of [15].

Let $a \subset \Omega_M$ be a subset of locations that can be simultaneously inspected, referred to as the inspection region hereafter, and let $\mathcal{A}_M = 2^{\Omega_M}$ be the collection of all allowable inspection regions. We assume that the outcome of an inspection depends on the size of the inspection region. More precisely, the outcome of inspecting region a , where $|a| = n$, is a random variable with Bernoulli distribution:

$$q_i^a = \begin{cases} \mathcal{B}(1 - p_n), & \text{if } i \in a \text{ and } |a| = n \\ \mathcal{B}(p_n), & \text{if } i \notin a \text{ and } |a| = n \end{cases} \quad \forall i \in \Omega_M, \forall a \in \mathcal{A}_M,$$

where $p_1 > 0$ and for all n , $p_n \leq p_{n+1}$ and $p_n \leq p$ for some $p < 0.5$.

LEMMA 3. *Consider the problem of a noisy dynamic search with size-dependent Bernoulli noise explained above. We have*

$$(22) \quad \min_{j \neq i} \max_{a \in \mathcal{A}_M} D(q_i^a \| q_j^a) = D_{\eta_i}(M) = (1 - 2p_1) \log \frac{1 - p_1}{p_1} \quad \forall i \in \Omega_M,$$

$$(23) \quad \underline{D}_2 = \overline{D}_{\max} = (1 - 2p_1) \log \frac{1 - p_1}{p_1},$$

$$(24) \quad 0 < 1 - \sup_n H([p_n, 1 - p_n]) \leq \underline{I}_2 \leq \overline{I}_{\max} \leq 1 - H([p_1, 1 - p_1]).$$

The proof is provided in the supplemental article [44], Section 4.

Lemma 3, together with Corollaries 2 and 4, implies that $\tilde{\pi}_2$ attains asymptotic optimality of order-2 in L , and order optimality in L and M . Furthermore, for the special case of size-independent Bernoulli noise where $0 < p_1 = p_2 = \dots = p < 0.5$, policy $\tilde{\pi}_2$ attains asymptotic optimality of order-1 in L and M .

The active hypothesis testing scheme proposed by Chernoff [17] as well as its variants [8, 11], when specialized to a noisy dynamic search with size-independent Bernoulli noise, simplifies to one that inspects, at each instant, a location with the highest probability of having the target. This scheme, which was also studied in [15] in a finite horizon context, has an information acquisition rate that is restricted to zero, while at zero rate, it achieves asymptotic optimality and maximum error exponent $\overline{D}_{\max} = (1 - 2p) \log \frac{1-p}{p}$. In contrast, in [14, 29], a noisy binary search was proposed in which the locations are partitioned along the median of the posterior and, in effect, are inspected along a generalized binary tree. It was shown in [14, 29] that the proposed policy can achieve any rate $R < 1 - H([p, 1 - p])$

¹¹Of course in this paper we are interested in a sequential setting where the sample size is not fixed a priori and is determined by the observation outcomes.

with reliability $E(R) = 1 - H([p, 1 - p]) - R$. In other words, the proposed policy in [14, 29] is asymptotically optimal in M (since $1 - H([p, 1 - p]) = \bar{I}_{\max}$) but only order optimal in L (since $0 < 1 - H([p, 1 - p]) < (1 - 2p) \log \frac{1-p}{p} = \bar{D}_{\max}$). Lemma 3 shows that, in the case of size-independent Bernoulli noise, our proposed policy $\tilde{\pi}_2$ combines the best of the above two approaches: in its first phase, by randomly selecting actions from \mathcal{A}_M , it ensures the maximum acquisition rate obtained by the noisy binary search of [14, 29], while its second phase coincides with the schemes in [8, 11, 17], ensuring the maximum feasible error exponent.

REMARK 6. Lemma 3 can be extended beyond the Bernoulli noise model so long as the observation kernels

$$q_i^a(\cdot) = \begin{cases} f_n(\cdot), & \text{if } i \in a \text{ and } |a| = n \\ \bar{f}_n(\cdot), & \text{if } i \notin a \text{ and } |a| = n \end{cases} \quad \forall i \in \Omega_M, \forall a \in \mathcal{A}_M,$$

satisfy the following conditions:

- (25) $f_n(z) = \bar{f}_n(b - z) \quad \forall z \in \mathcal{Z} \text{ for some } b \in \mathbb{R},$
- (26) $D(f_n \|\alpha f_n + \bar{\alpha} \bar{f}_n) \geq D(f_{n+1} \|\alpha f_{n+1} + \bar{\alpha} \bar{f}_{n+1}) \quad \forall \alpha \in [0, 1], \bar{\alpha} = 1 - \alpha,$
- (27) $\sup_n \sup_{z \in \mathcal{Z}} f_n(z) / \bar{f}_n(z) < \infty.$

In particular, under these conditions

- (28) $\min_{j \neq i} \max_{a \in \mathcal{A}_M} D(q_i^a \| q_j^a) = D_{\eta_i}(M) = D(f_1 \| \bar{f}_1) \quad \forall i \in \Omega_M,$
- (29) $\underline{D}_2 = \bar{D}_{\max} = D(f_1 \| \bar{f}_1),$
- (30) $\inf_n D\left(f_n \left\| \frac{1}{2} f_n + \frac{1}{2} \bar{f}_n\right.\right) \leq \underline{L}_2 \leq \bar{I}_{\max} \leq D\left(f_1 \left\| \frac{1}{2} f_1 + \frac{1}{2} \bar{f}_1\right.\right).$

Condition (25) implies that given a fixed inspection area, the collected samples provide identical information regarding the presence of the target or its absence. Condition (26) implies that the samples become less informative as the size of the inspection region increases. Conditions (25) and (26) are natural, while condition (27) is a technical one to ensure that Assumptions 2 and 3 hold (we address weakening these assumptions in Section 7).

7. Discussions. In this section we provide a discussion on the technical assumptions of the paper. In particular, we discuss the necessity of our Assumptions 1 and 2, and compare them with the common assumptions in the literature. In contrast to Assumption 1 which is shown to be necessary for the problem of active hypothesis testing to have a meaningful solution, Assumption 2 can be relaxed to more general assumptions without affecting the asymptotic results of the paper.

7.1. *Assumption 1.* We first discuss the necessity of Assumption 1. If Assumption 1 does not hold, then there exist two hypotheses $i, j \in \Omega_M$, $i \neq j$ such that for all $a \in \mathcal{A}_M$, $D(q_i^a \| q_j^a) = 0$. In other words, $q_i^a(\cdot) = q_j^a(\cdot)$ for all $a \in \mathcal{A}_M$, and, hence, the decision maker is not capable of distinguishing these two hypotheses. In this sense, Assumption 1 is necessary for Problem (P) to be meaningful.

Next we compare Assumption 1 to its counterpart in [17]:

ASSUMPTION 1'. $D(q_i^a \| q_j^a) > 0, \forall i, j \in \Omega_M, i \neq j, \forall a \in \mathcal{A}_M$.

This assumption assures *consistency* (see Lemma 1 in [17]), that is, $\arg \max_{i \in \Omega_M} \rho_i(t)$ converges exponentially fast to the true hypothesis regardless of the way the sensing actions are selected. However, this assumption is very restrictive and does not hold in many problems of interest such as channel coding with feedback [12] and noisy dynamic search (e.g., one cannot discriminate between locations 1 and 2 by inspecting location 3). It was remarked in [17], Section 7, that the above restrictive assumption can be relaxed if the proposed scheme is modified to take a (possibly randomized) action capable of discriminating between all hypotheses pairs infinitely often (e.g., at any time t when t is a perfect square). In this paper, however, we took a different approach and constructed policy $\tilde{\pi}_1$, a simple two-phase modification of Chernoff's original scheme in which testing for the maximum likely hypothesis is delayed and contingent on obtaining a certain level of confidence.

7.2. *Assumption 2.* We first discuss the necessity of Assumption 2. For observation kernels with bounded support, Assumption 2 is a necessary condition to ensure that the observation kernels are absolutely continuous with respect to each other and, hence, no observation is noise free. Although this assumption might hold in many settings such as the problem of a noisy dynamic search with Bernoulli noise explained in Section 6.2, it does not hold in general for observation kernels with unbounded support such as Gaussian distribution. Next we replace Assumption 2 by more general assumptions on the observation kernels and discuss the consequences.

To the best of our knowledge, Assumption 2' below, given first by [17], is the weakest condition in the literature of hypothesis testing and sequential analysis, and is often interpreted to an assumption which limits the *excess over the boundary* at the stopping time [38].

ASSUMPTION 2'. There exists $\xi_M < \infty$ such that

$$\max_{i, j \in \Omega_M} \max_{a \in \mathcal{A}_M} \int_{\mathcal{Z}} q_i^a(z) \left| \log \frac{q_i^a(z)}{q_j^a(z)} \right|^2 dz \leq \xi_M.$$

Proposition 1 remains valid even if Assumption 2 is replaced with Assumption 2' (with the only change that ξ_M^2 is replaced with ξ_M in the bound). The proof of Proposition 2 relies on Chernoff's approach [17], and the asymptotic behavior of the bound remains intact if Assumption 2 is replaced with Assumption 2'. However, as shown in the proof of this proposition in the supplemental article [44], Section 5, Assumption 2 allows us to give a precise nonasymptotic characterization of the bound by applying the *method of bounded differences* and, in particular, McDiarmid's inequality [41].

Next we consider the consequence of weakening Assumption 2 on Theorems 2 and 3, hence on the performance of policy $\tilde{\pi}_2$. To do so, we consider an even weaker assumption than Assumption 2' as given below:

ASSUMPTION 2''. There exist $\xi_M < \infty$ and $\gamma > 0$ such that

$$\max_{i,j \in \Omega_M} \max_{a \in \mathcal{A}_M} \int_{\mathcal{Z}} q_i^a(z) \left| \log \frac{q_i^a(z)}{q_j^a(z)} \right|^{1+\gamma} dz \leq \xi_M.$$

Define function $\psi_M : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ as follows:

$$\psi_M(b) := \max_{i,j \in \Omega_M} \max_{a \in \mathcal{A}_M} \int_{\mathcal{Z}} q_i^a(z) \left[\log \frac{q_i^a(z)}{q_j^a(z)} \right]_b dz,$$

where $[g]_b = g \mathbf{1}_{\{g > b\}}$. Note that $\psi_M(b)$ is in general nonincreasing in b , and if Assumption 2'' holds, $\psi_M(b) \leq b^{-\gamma} \xi_M$. Under the weaker Assumption 2'' (and naturally Assumption 2'), Theorems 2 and 3 can be replaced by the following:

PROPOSITION 3. Under Assumptions 1 and 2'' and for $L > \frac{\log M}{I_{\max}(M)}$, $\rho \in \mathbb{P}_L(\Omega_M)$, $\delta \in (0, 0.5]$, and $b > 0$,

$$\begin{aligned} V^*(\rho) \geq \underline{V}_3(\rho) := & \frac{1}{1 + \psi_M(b)/D_{\max}(M)} \\ & \times \left[\frac{H(\rho) - H([\delta, 1 - \delta]) - \delta \log(M - 1)}{I_{\max}(M)} \right. \\ & \left. + \frac{\log((1 - L^{-1})/L^{-1}) - \log((1 - \delta)/\delta) - b}{D_{\max}(M)} \right. \\ & \left. \times \mathbf{1}_{\{\max_{i \in \Omega_M} \rho_i \leq 1 - \delta\}} - K'_3 \right]^+, \end{aligned}$$

where K'_3 is a positive constant independent of δ and L . In addition, if Assumption 3 also holds, then K'_3 can be selected independent of M as well.

The proof is provided in the supplemental article [44], Section 8.1.

PROPOSITION 4. Under Assumptions 1 and 2'', and for $L > 1$ and $\rho \in \mathbb{P}_L(\Omega_M)$, $\exists b' \in (0, \infty)$ such that for all $b \geq b'$, $0 \leq \frac{(1+(\log e)/b)2^{-b}\psi_M(b)}{I_2(M)-\psi_M(b)} < 1$, and

$$\begin{aligned}
 V^*(\rho) &\leq \bar{V}_3(\rho) \\
 &:= \left(1 - \frac{(1 + (\log e)/b)2^{-b}\psi_M(b)}{I_2(M) - \psi_M(b)}\right)^{-1} \\
 &\quad \times \left(\frac{H(\rho) + \log(\tilde{\rho}/(1 - \tilde{\rho})) + \bar{b} + \log e}{I_2(M) - \psi_M(b)} + \sum_{i=1}^M \rho_i \frac{\log L}{D_{\eta_i}(M) - \psi_M(b)}\right) \\
 &\quad + 1.
 \end{aligned}$$

The proof is provided in the supplemental article [44], Section 8.2.

As we discussed, $\psi_M(b) \leq b^{-\gamma}\xi_M$ under Assumption 2''. Furthermore, if Assumption 3 holds, then $\sup_M \psi_M(b) \leq b^{-\gamma}\xi$. In other words, we can select b as a function of L and M (e.g., $b = \log \log LM$) such that \underline{V}_3 and \bar{V}_3 have the same dominating terms (in L and M) as \underline{V}_2 and \bar{V}_2 , respectively.

In summary, the asymptotic results of the paper presented in Section 5 hold under the weaker Assumptions 2' and 2'' replacing Assumption 2 (with the only exception that the asymptotic optimality of order-2 of policy $\tilde{\pi}_2$ established in Corollary 2 is degraded to asymptotic optimality of order-1). Our choice to present the work under Assumption 2, however, significantly simplifies the presentation and also enables a precise nonasymptotic characterization of the lower and upper bounds.

8. Conclusions and future work. In this paper we considered the problem of active sequential M -ary hypothesis testing. Using a DP formulation, we characterized the optimal value function V^* . Three lower bounds (complementary for various values of the parameters of the problem) were obtained for the optimal value function V^* . We also proposed two heuristic policies whose performance analysis resulted in two upper bounds for V^* . Subsequently, we discussed important consequences of the bounds and established order and asymptotic optimality of the proposed policies under different scenarios. An important problem which remains is further improvement of the performance bounds.

In this paper we focused on sequential policies, that is, policies whose sample size is not known initially and is dependent on the observation outcomes. There exist other types of policies in the literature. For example, nonsequential policies take a fixed number of samples (independent of observation outcomes) and make the final decision afterward, while multi-stage policies (introduced in [3, 40]) can take a retire-declare action only at the end of each stage, and stages are not necessarily of the same size. Comparing the performance of sequential, nonsequential and multi-stage policies in the context of active hypothesis testing is an area of future work.

In this paper we assumed that all sensing actions incur one unit of cost (each action can be executed in one unit of time). It is also of interest to consider the scenario where there is a cost associated with each action which, for example, characterizes the amount of energy or time required to perform that action; and the goal is to find the true hypothesis subject to a cost criterion. Such generalization has been studied for the problem of variable-length coding with feedback in [45].

APPENDIX: PROOF OF THEOREMS 1–3

A.1. Proof of Theorem 1. Let Γ be the set of all mappings $\gamma : \Omega_M \rightarrow \Omega_M$ such that $\gamma(i) \neq i$ for $i \in \Omega_M$. Now associated with any $\gamma \in \Gamma$, define

$$(31) \quad \underline{V}_1^\gamma(\rho) = \left[\sum_{i=1}^M \rho_i \frac{\log((1 - L^{-1})/L^{-1}) - \log(\rho_i/\rho_{\gamma(i)})}{\max_{\hat{a} \in \mathcal{A}_M} D(q_i^{\hat{a}} \| q_{\gamma(i)}^{\hat{a}})} - K'_1 \right]^+.$$

Next we use Lemma 1 to show that $V^* \geq \underline{V}_1^\gamma$ for all $\gamma \in \Gamma$. In particular, we show that for all $\gamma \in \Gamma$ and all $\rho \in \mathbb{P}(\Omega_M)$, $\underline{V}_1^\gamma(\rho) \leq \min\{1 + \min_{a \in \mathcal{A}_M} (\mathbb{T}^a \underline{V}_1^\gamma)(\rho), \min_{j \in \Omega_M} (1 - \rho_j)L\}$. For any ρ such that $\underline{V}_1^\gamma(\rho) = 0$, the inequality holds trivially. For $\underline{V}_1^\gamma(\rho) > 0$ and for any action $a \in \mathcal{A}_M$, we have

$$\begin{aligned} & (\mathbb{T}^a \underline{V}_1^\gamma)(\rho) \\ & \geq \sum_{i=1}^M \int \rho_i q_i^a(z) \frac{\log((1 - L^{-1})/L^{-1}) - \log(\rho_i q_i^a(z)/(\rho_{\gamma(i)} q_{\gamma(i)}^a(z)))}{\max_{\hat{a} \in \mathcal{A}_M} D(q_i^{\hat{a}} \| q_{\gamma(i)}^{\hat{a}})} dz \\ & \quad - K'_1 \\ & = \underline{V}_1^\gamma(\rho) - \sum_{i=1}^M \rho_i \frac{D(q_i^a \| q_{\gamma(i)}^a)}{\max_{\hat{a} \in \mathcal{A}_M} D(q_i^{\hat{a}} \| q_{\gamma(i)}^{\hat{a}})} \\ & \geq \underline{V}_1^\gamma(\rho) - 1. \end{aligned}$$

CLAIM 1 (In Section 9.1 of the supplemental article [44]). *Constant K'_1 can be selected independent of L such that $\underline{V}_1^\gamma(\rho) \leq \min_{j \in \Omega_M} (1 - \rho_j)L$ is satisfied for all $\gamma \in \Gamma$.*

Using Claim 1 and letting $\underline{V}_1(\cdot) = \max_{\gamma \in \Gamma} \underline{V}_1^\gamma(\cdot)$, we have the assertion of the theorem.

A.2. Proof of Theorem 2. We first show that for all $\rho \in \mathbb{P}(\Omega_M)$,

$$(32) \quad V^*(\rho) \geq \left[\frac{H(\rho) - H([\alpha(L, M), 1 - \alpha(L, M)]) - \alpha(L, M) \log(M - 1)}{I_{\max}(M)} + \alpha(L, M)L \right]^+.$$

Note that the right-hand side of (32) can be written as

$$(33) \quad G(\boldsymbol{\rho}) := \left[\frac{H(\boldsymbol{\rho}) - H(\mathbf{v})}{I_{\max}(M)} + \alpha(L, M)L \right]^+,$$

where

$$(34) \quad \mathbf{v} = \left[\frac{\alpha(L, M)}{M - 1}, \dots, \frac{\alpha(L, M)}{M - 1}, 1 - \alpha(L, M) \right].$$

Next we show that $G(\boldsymbol{\rho}) \leq \min\{1 + \min_{a \in \mathcal{A}_M} (\mathbb{T}^a G)(\boldsymbol{\rho}), \min_{j \in \Omega_M} (1 - \rho_j)L\}$ for all $\boldsymbol{\rho} \in \mathbb{P}(\Omega_M)$. For any $\boldsymbol{\rho}$ such that $G(\boldsymbol{\rho}) = 0$, the inequality holds trivially. For $G(\boldsymbol{\rho}) > 0$ and for any action $a \in \mathcal{A}_M$, we have

$$(35) \quad \begin{aligned} (\mathbb{T}^a G)(\boldsymbol{\rho}) &= \frac{\int H(\Phi^a(\boldsymbol{\rho}, z))q_{\boldsymbol{\rho}}^a(z) dz - H(\mathbf{v})}{I_{\max}(M)} + \alpha(L, M)L \\ &= \frac{H(\boldsymbol{\rho}) - I(\boldsymbol{\rho}; q_{\boldsymbol{\rho}}^a) - H(\mathbf{v})}{I_{\max}(M)} + \alpha(L, M)L \\ &= G(\boldsymbol{\rho}) - \frac{I(\boldsymbol{\rho}; q_{\boldsymbol{\rho}}^a)}{I_{\max}(M)} \\ &\geq G(\boldsymbol{\rho}) - 1, \end{aligned}$$

where the last inequality follows from the fact that

$$I(\boldsymbol{\rho}; q_{\boldsymbol{\rho}}^a) \leq \max_{\hat{a} \in \mathcal{A}_M} \max_{\hat{\boldsymbol{\rho}} \in \mathbb{P}(\Omega_M)} I(\hat{\boldsymbol{\rho}}; q_{\hat{\boldsymbol{\rho}}}^{\hat{a}}) = I_{\max}(M).$$

Therefore,

$$G(\boldsymbol{\rho}) \leq 1 + \min_{a \in \mathcal{A}_M} (\mathbb{T}^a G)(\boldsymbol{\rho}).$$

What remains is to show that $G(\boldsymbol{\rho}) \leq \min_{j \in \Omega_M} (1 - \rho_j)L$. Rewriting G as

$$G(\boldsymbol{\rho}) = \left[\frac{\sum_{i=1}^{M-1} \rho_i \log(1/\rho_i) + (1 - \sum_{i=1}^{M-1} \rho_i) \log(1/(1 - \sum_{i=1}^{M-1} \rho_i)) - H(\mathbf{v})}{I_{\max}(M)} + \alpha(L, M)L \right]^+,$$

we can compute the gradient at \mathbf{v} . For all $i = 1, 2, \dots, M - 1$,

$$\begin{aligned} \frac{\partial G}{\partial \rho_i}(\mathbf{v}) &= \left(\log \frac{1}{\rho_i} - \log e - \log \frac{1}{1 - \sum_{i=1}^{M-1} \rho_i} + \log e \right) / I_{\max}(M) \Big|_{\boldsymbol{\rho}=\mathbf{v}} \\ &= \left(\log \frac{\rho_M}{\rho_i} \right) / I_{\max}(M) \Big|_{\boldsymbol{\rho}=\mathbf{v}} = \left(\log \frac{1 - \alpha(L, M)}{\alpha(L, M)/(M - 1)} \right) / I_{\max}(M) = L. \end{aligned}$$

Furthermore, $G(\mathbf{v}) = \alpha(L, M)L = (1 - \nu_M)L$. Without loss of generality and since both functions $G(\boldsymbol{\rho})$ and $\min_{j \in \Omega_M} (1 - \rho_j)L$ are symmetric, let us focus on $\mathbb{P}_M(\Omega_M) := \{\boldsymbol{\rho} \in \mathbb{P}(\Omega_M) : \rho_M \geq \rho_i, \forall i \in \Omega_M - \{M\}\}$. In this case, $\min_{j \in \Omega_M} (1 - \rho_j)L = (1 - \rho_M)L = \sum_{i=1}^{M-1} \rho_i L$ and, hence, $\min_{j \in \Omega_M} (1 - \rho_j)L$

is the tangent hyperplane to $G(\boldsymbol{\rho})$ at \mathbf{v} . This along with concavity of function G implies $G(\boldsymbol{\rho}) \leq \min_{j \in \Omega_M} (1 - \rho_j)L$. Using Lemma 1, we have the assertion of the theorem.

Next we need to show that

$$(36) \quad V^*(\boldsymbol{\rho}) \geq \underline{V}_2(\boldsymbol{\rho}) = \left[\frac{H(\boldsymbol{\rho}) - H([\delta, 1 - \delta]) - \delta \log(M - 1)}{I_{\max}(M)} + \frac{\log((1 - L^{-1})/L^{-1}) - \log((1 - \delta)/\delta) - \xi_M}{D_{\max}(M)} \times \mathbf{1}_{\{\max_{i \in \Omega_M} \rho_i \leq 1 - \delta\}} - K'_2 \right]^+.$$

We show this in two steps. First we consider the following function:

$$(37) \quad J'(\boldsymbol{\rho}) := \left[\sum_{i=1}^M \rho_i \frac{\log((1 - L^{-1})/L^{-1}) - \log(\rho_i/(1 - \rho_i))}{D_{\max}(M)} - K'_2 \right]^+.$$

We use Jensen’s inequality to show that

$$(38) \quad J'(\boldsymbol{\rho}) \leq 1 + \min_{a \in \mathcal{A}_M} (\mathbb{T}^a J')(\boldsymbol{\rho}) \quad \forall \boldsymbol{\rho} \in \mathbb{P}(\Omega_M).$$

For any $\boldsymbol{\rho}$ such that $J'(\boldsymbol{\rho}) = 0$, inequality (38) holds trivially. For any $\boldsymbol{\rho}$ such that $J'(\boldsymbol{\rho}) > 0$ and for any $a \in \mathcal{A}_M$, we have

$$\begin{aligned} & (\mathbb{T}^a J')(\boldsymbol{\rho}) \\ & \geq \sum_{i=1}^M \int \rho_i q_i^a(z) \frac{\log((1 - L^{-1})/L^{-1}) - \log(\rho_i q_i^a(z) / \sum_{j \neq i} \rho_j q_j^a(z))}{D_{\max}(M)} dz \\ & \quad - K'_2 \\ & = J'(\boldsymbol{\rho}) - \sum_{i=1}^M \rho_i \frac{\int q_i^a(z) \log(q_i^a(z) / \sum_{j \neq i} (\rho_j / (1 - \rho_i)) q_j^a(z)) dz}{D_{\max}(M)} \\ & \geq J'(\boldsymbol{\rho}) - \sum_{i=1}^M \rho_i \frac{\sum_{j \neq i} (\rho_j / (1 - \rho_i)) D(q_i^a \| q_j^a)}{D_{\max}(M)} \\ & \geq J'(\boldsymbol{\rho}) - 1. \end{aligned}$$

Next we define $J(\boldsymbol{\rho}) = \max\{J'(\boldsymbol{\rho}), J''(\boldsymbol{\rho})\}$, where $J''(\boldsymbol{\rho})$ is the right-hand side of (36), that is,

$$(39) \quad J''(\boldsymbol{\rho}) = \left[\frac{H(\boldsymbol{\rho}) - H([\delta, 1 - \delta]) - \delta \log(M - 1)}{I_{\max}(M)} + \frac{\log((1 - L^{-1})/L^{-1}) - \log((1 - \delta)/\delta) - \xi_M}{D_{\max}(M)} \times \mathbf{1}_{\{\max_{i \in \Omega_M} \rho_i \leq 1 - \delta\}} - K'_2 \right]^+.$$

- *Case 1:* For all ρ such that $J(\rho) = 0$ or $J(\rho) = J'(\rho)$, it is trivial from (38) that

$$(40) \quad J(\rho) = J'(\rho) \leq 1 + \min_{a \in \mathcal{A}_M} (\mathbb{T}^a J')(\rho) \leq 1 + \min_{a \in \mathcal{A}_M} (\mathbb{T}^a J)(\rho).$$

- *Case 2:* For all ρ such that $J(\rho) = J''(\rho) > 0$, and for any action $a \in \mathcal{A}_M$, we have

$$\begin{aligned} (\mathbb{T}^a J)(\rho) &= \int J(\Phi^a(\rho, z))q_\rho^a(z) dz \\ &\stackrel{(a)}{\geq} \frac{\int H(\Phi^a(\rho, z))q_\rho^a(z) dz - H([\delta, 1 - \delta]) - \delta \log(M - 1)}{I_{\max}(M)} \\ &\quad + \frac{\log((1 - L^{-1})/L^{-1}) - \log((1 - \delta)/\delta) - \xi_M}{D_{\max}(M)} \\ (41) \quad &\quad \times \mathbf{1}_{\{\max_{i \in \Omega_M} \rho_i \leq 1 - \delta\}} - K'_2 \\ &= J''(\rho) - \frac{I(\rho; q_\rho^a)}{I_{\max}(M)} \\ &\geq J''(\rho) - 1 \\ &\stackrel{(b)}{=} J(\rho) - 1, \end{aligned}$$

where (a) follows from Claim 2 below and (b) holds since ρ is such that $J(\rho) = J''(\rho)$.

CLAIM 2 (In Section 9.2 of the supplemental article [44]). *Let ρ be such that $J(\rho) = J''(\rho) > 0$. If Assumption 2 holds, then for all actions $a \in \mathcal{A}_M$ and observations $z \in \mathcal{Z}$,*

$$\begin{aligned} J(\Phi^a(\rho, z)) &\geq \frac{H(\Phi^a(\rho, z)) - H([\delta, 1 - \delta]) - \delta \log(M - 1)}{I_{\max}(M)} \\ (42) \quad &\quad + \frac{\log((1 - L^{-1})/L^{-1}) - \log((1 - \delta)/\delta) - \xi_M}{D_{\max}(M)} \\ &\quad \times \mathbf{1}_{\{\max_{i \in \Omega_M} \rho_i \leq 1 - \delta\}} - K'_2. \end{aligned}$$

Combining (40) and (41), we have that

$$(43) \quad J(\rho) \leq 1 + \min_{a \in \mathcal{A}_M} (\mathbb{T}^a J)(\rho).$$

We also have the following:

CLAIM 3 (In Section 9.3 of the supplemental article [44]). *For $L > \frac{\log M}{I_{\max}(M)}$, constant K'_2 can be selected independent of δ and L such that $J(\rho) \leq \min_{j \in \Omega_M} (1 - \rho_j)L$. Furthermore, if $\sup_M \xi_M < \infty$, then K'_2 can be selected independent of M as well.*

Lemma 1, together with (43) and Claim 3, implies that $V^* \geq J = \max\{J', J''\} \geq J'' = \underline{V}_2$. This is a slightly stronger result than (36).

A.3. Proof of Theorem 3. Recall that $\rho_i(n)$ denotes the posterior belief about hypothesis H_i after n observations. Let $\tau, \tau_i, i \in \Omega_M$, be Markov stopping times defined as follows:

$$(44) \quad \tau := \min\left\{n : \max_{j \in \Omega_M} \rho_j(n) \geq 1 - L^{-1}\right\},$$

$$(45) \quad \tau_i := \min\{n : \rho_i(n) \geq 1 - L^{-1}\}.$$

From (1), the expected total cost under policy $\tilde{\pi}_2$ is upper bounded as

$$(46) \quad \begin{aligned} V_{\tilde{\pi}_2}(\boldsymbol{\rho}) &= \mathbb{E}_{\tilde{\pi}_2} \left[\tau + \min_{j \in \Omega_M} (1 - \rho_j(\tau))L \right] \\ &\leq \mathbb{E}_{\tilde{\pi}_2}[\tau] + 1 \\ &\leq \sum_{i=1}^M \rho_i \mathbb{E}_{\tilde{\pi}_2}[\tau_i | \theta = i] + 1, \end{aligned}$$

where $\boldsymbol{\rho} = [\rho_1, \rho_2, \dots, \rho_M] = [\rho_1(0), \rho_2(0), \dots, \rho_M(0)]$ and the last inequality follows from the fact that $\tau \leq \tau_i, \forall i \in \Omega_M$. For notational simplicity, subscript $\tilde{\pi}_2$ is dropped for the rest of the proof.

Next we find an upper bound for $\mathbb{E}[\tau_i | \theta = i], i \in \Omega_M$. Let

$$(47) \quad U_n := \log \frac{\rho_i(n)}{1 - \rho_i(n)} - \log \frac{\tilde{\rho}}{1 - \tilde{\rho}}$$

and let \mathcal{F}_n denote the history of previous actions and observations up to time n , that is, $\mathcal{F}_n := \sigma\{\boldsymbol{\rho}(0), A(0), Z(0), \dots, A(n - 1), Z(n - 1)\}$. Under policy $\tilde{\pi}_2$, the sequence $\{U_n\}, n = 0, 1, \dots$, forms a submartingale with respect to the filtration $\{\mathcal{F}_n\}$ with the following properties:

(C1) If $U_n < 0$ and $\rho_j(n) < \tilde{\rho}$ for all $j \in \Omega_M (\Rightarrow P(A(n) = a) = \eta_{0a})$:

$$\begin{aligned} &\mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i] \\ &= \sum_{a \in \mathcal{A}_M} P(A(n) = a) \mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i, A(n) = a] \\ &= \sum_{a \in \mathcal{A}_M} \eta_{0a} \mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i, A(n) = a] \\ &= \sum_{a \in \mathcal{A}_M} \eta_{0a} \mathbb{E} \left[\log \frac{\rho_i(n) q_i^a(Z)}{\sum_{j \neq i} \rho_j(n) q_j^a(Z)} - \log \frac{\rho_i(n)}{1 - \rho_i(n)} \middle| \mathcal{F}_n, \theta = i \right] \\ &= \sum_{a \in \mathcal{A}_M} \eta_{0a} \int q_i^a(z) \log \frac{q_i^a(z)}{\sum_{j \neq i} (\rho_j(n)/(1 - \rho_i(n))) q_j^a(z)} dz \end{aligned}$$

$$\begin{aligned} &\geq \max_{\lambda \in \mathbb{P}(\mathcal{A}_M)} \min_{i \in \Omega_M} \min_{\hat{\rho} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a D\left(q_i^a \parallel \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a\right) \\ &= I_{\eta_0}(M). \end{aligned}$$

If $U_n < 0$ and $\rho_k(n) \geq \tilde{\rho}$ for some $k \neq i$ ($\Rightarrow P(A(n) = a) = \eta_{ka}$):

$$\begin{aligned} &\mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i] \\ &= \sum_{a \in \mathcal{A}_M} \eta_{ka} \mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i, A(n) = a] \\ &= \sum_{a \in \mathcal{A}_M} \eta_{ka} \int q_i^a(z) \log \frac{q_i^a(z)}{\sum_{j \neq i} (\rho_j(n)/(1 - \rho_i(n))) q_j^a(z)} dz \\ &\geq \min_{i \in \Omega_M} \min_{k \neq i} \min_{\hat{\rho} : \hat{\rho}_k \geq \tilde{\rho}} \sum_{a \in \mathcal{A}_M} \eta_{ka} D\left(q_i^a \parallel \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a\right) \\ &= I_{\eta, \tilde{\rho}}(M); \end{aligned}$$

(C2) If $U_n \geq 0$ ($\rho_i(n) \geq \tilde{\rho} \Rightarrow P(A(n) = a) = \eta_{ia}$):

$$\begin{aligned} &\mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i] \\ &= \sum_{a \in \mathcal{A}_M} \eta_{ia} \mathbb{E}[U_{n+1} - U_n | \mathcal{F}_n, \theta = i, A(n) = a] \\ &= \sum_{a \in \mathcal{A}_M} \eta_{ia} \int q_i^a(z) \log \frac{q_i^a(z)}{\sum_{j \neq i} (\rho_j(n)/(1 - \rho_i(n))) q_j^a(z)} dz \\ &\geq \max_{\lambda \in \mathbb{P}(\mathcal{A}_M)} \min_{\hat{\rho} \in \mathbb{P}_L(\Omega_M)} \sum_{a \in \mathcal{A}_M} \lambda_a D\left(q_i^a \parallel \sum_{j \neq i} \frac{\hat{\rho}_j}{1 - \hat{\rho}_i} q_j^a\right) \\ &= D_{\eta_i}(M); \end{aligned}$$

(C3) $|U_n - U_{n-1}| \leq \max_{i, j \in \Omega_M} \max_{a \in \mathcal{A}_M} \sup_{z \in \mathcal{Z}} \log \frac{q_i^a(z)}{q_j^a(z)} \leq \xi_M$.

Stopping time τ_i defined in (45) can be rewritten as

$$\begin{aligned} &\tau_i = \min\{n : \rho_i(n) \geq 1 - L^{-1}\} \\ &= \min\left\{n : \frac{\rho_i(n)}{1 - \rho_i(n)} \geq \frac{1 - L^{-1}}{L^{-1}}\right\} \\ (48) \quad &= \min\left\{n : \log \frac{\rho_i(n)}{1 - \rho_i(n)} - \log \frac{\tilde{\rho}}{1 - \tilde{\rho}} \geq \log \frac{1 - L^{-1}}{L^{-1}} - \log \frac{\tilde{\rho}}{1 - \tilde{\rho}}\right\} \\ &= \min\left\{n : U_n \geq \log \frac{1 - L^{-1}}{L^{-1}} - \log \frac{\tilde{\rho}}{1 - \tilde{\rho}}\right\} \\ &\leq \min\{n : U_n \geq \log L\}. \end{aligned}$$

The assertion of the theorem follows from (48) and the following lemma.

LEMMA 4. Consider the sequence $\{U_n\}$, $n = 0, 1, \dots$ defined in (47), and assume there exist positive constants $K_1 \leq K_2 \leq K_3$ such that

$$\begin{aligned} \mathbb{E}[U_{n+1}|\mathcal{F}_n, \theta = i] &\geq U_n + K_1 && \text{if } U_n < 0, \\ \mathbb{E}[U_{n+1}|\mathcal{F}_n, \theta = i] &\geq U_n + K_2 && \text{if } U_n \geq 0, \\ |U_{n+1} - U_n| &\leq K_3. \end{aligned}$$

Consider the stopping time $\nu = \min\{n : U_n \geq B\}$, $B > [U_0]^+$. Then we have

$$\mathbb{E}[\nu|\theta = i] \leq \frac{B - U_0}{K_2} + U_0 \mathbf{1}_{\{U_0 < 0\}} \left(\frac{1}{K_2} - \frac{1}{K_1} \right) + \frac{K_3 + \log e}{K_1}.$$

The proof of Lemma 4 is provided in the supplemental article [44], Section 6. In particular, from (C1)–(C3) and Lemma 4, we have

$$\begin{aligned} &\rho_i \mathbb{E}[\tau_i|\theta = i] \\ &\leq \rho_i \left(\frac{\log L - [\log(\rho_i/(1 - \rho_i)) - \log(\tilde{\rho}/(1 - \tilde{\rho}))]^+}{D_{\eta_i}(M)} \right. \\ &\quad \left. + \frac{[\log((1 - \rho_i)/\rho_i) + \log(\tilde{\rho}/(1 - \tilde{\rho}))]^+ + \xi_M + \log e}{I_2(M)} \right) \\ &\leq \rho_i \frac{\log L}{D_{\eta_i}(M)} + \rho_i \frac{\log(1/\rho_i) + \log(\tilde{\rho}/(1 - \tilde{\rho})) + \xi_M + \log e}{I_2(M)}. \end{aligned}$$

This inequality together with (46) and the fact that $\sum_{i=1}^M \rho_i \log \frac{1}{\rho_i} = H(\rho)$ implies the assertion of the theorem:

$$\begin{aligned} (49) \quad V^*(\rho) &\leq V_{\tilde{\pi}_2}(\rho) \\ &\leq \frac{H(\rho) + \log(\tilde{\rho}/(1 - \tilde{\rho})) + \xi_M + \log e}{I_2(M)} + \sum_{i=1}^M \rho_i \frac{\log L}{D_{\eta_i}(M)} + 1. \end{aligned}$$

REMARK 7. For large values of $H(\rho)$ and $\tilde{\rho}$ and when $I_{\eta_0}(M) > I_{\eta, \tilde{\rho}}(M)$, the upper bound (49) can be tightened as follows (see Section 7 in [44] for the proof):

$$\begin{aligned} (50) \quad V_{\tilde{\pi}_2}(\rho) &\leq \frac{H(\rho) + \log(\tilde{\rho}/(1 - \tilde{\rho})) + \xi_M}{I_{\eta_0}(M)} + \sum_{i=1}^M \rho_i \frac{\log L}{D_{\eta_i}(M)} \\ &\quad + \frac{(1 - \tilde{\rho}) \log M + (2 - \tilde{\rho}) \xi_M + 4 + \log e}{I_{\eta, \tilde{\rho}}(M)} + 1. \end{aligned}$$

Acknowledgments. We would like to thank Todd Coleman, Young-Han Kim, Barış Nakiboğlu, Yury Polyanskiy, Maxim Raginsky, Sergio Verdú, Michèle Wigger and Angela Yu for valuable discussions and suggestions. We are also grateful to the Editor, the Associate Editor and two reviewers for their constructive comments. This work was done while Mohammad Naghshvar was with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093, USA.

SUPPLEMENTARY MATERIAL

Technical proofs (DOI: [10.1214/13-AOS1144SUPP](https://doi.org/10.1214/13-AOS1144SUPP); .pdf). For the interest of space, we only provided the proofs of the theorems in this paper. Proofs of the propositions, lemmas, corollaries and technical claims are provided in the supplemental article.

REFERENCES

- [1] ALBERT, A. E. (1961). The sequential design of experiments for infinitely many states of nature. *Ann. Math. Statist.* **32** 774–799. [MR0125720](#)
- [2] ARMITAGE, P. (1950). Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **12** 137–144. [MR0038618](#)
- [3] BARTROFF, J. (2007). Asymptotically optimal multistage tests of simple hypotheses. *Ann. Statist.* **35** 2075–2105. [MR2363964](#)
- [4] BERLEKAMP, E. R. (1964). Block coding with noiseless feedback. Ph.D. thesis, MIT, Cambridge, MA.
- [5] BERLIN, P., NAKIBOĞLU, B., RIMOLDI, B. and TELATAR, E. (2009). A simple converse of Burnashev’s reliability function. *IEEE Trans. Inform. Theory* **55** 3074–3080. [MR2598006](#)
- [6] BERRY, S. M., CARLIN, B. P., LEE, J. J. and MÜLLER, P. (2011). *Bayesian Adaptive Methods for Clinical Trials*. Chapman & Hall/CRC Biostatistics Series **38**. CRC Press, Boca Raton, FL. [MR2723582](#)
- [7] BERTSEKAS, D. P. and SHREVE, S. E. (2007). *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific, Belmont, CA.
- [8] BESSLER, S. A. (1960). Theory and applications of the sequential design of experiments, K -actions and infinitely many experiments: Part I—Theory. Technical Report no. 55, Dept. Statistics, Univ. Stanford, Stanford, CA.
- [9] BLACKWELL, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist.* **24** 265–272. [MR0056251](#)
- [10] BLAHUT, R. E. (1974). Hypothesis testing and information theory. *IEEE Trans. Inform. Theory* **20** 405–417. [MR0396072](#)
- [11] BLOT, W. J. and MEETER, D. A. (1973). Sequential experimental design procedures. *J. Amer. Statist. Assoc.* **68** 586–593. [MR0359209](#)
- [12] BURNASHEV, M. V. (1975). Data transmission over a discrete channel with feedback. Random transmission time. *Problemy Peredachi Informatsii* **12** 10–30.
- [13] BURNASHEV, M. V. (1980). Sequential discrimination of hypotheses with control of observations. *Mathematics of the USSR–Izvestiya* **15** 419–440.
- [14] BURNASHEV, M. V. and ZIGANGIROV, K. S. (1974). A certain problem of interval estimation in observation control. *Problemy Peredachi Informatsii* **10** 51–61. [MR0469505](#)
- [15] CASTANON, D. A. (1995). Optimal search strategies in dynamic hypothesis testing. *IEEE Trans. Syst., Man Cybern.* **25** 1130–1138.

- [16] CHAN, C. L., CHE, P. H., JAGGI, S. and SALIGRAMA, V. (2011). Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms. In *49th Annual Allerton Conference on Communication, Control, and Computing* 1832–1839.
- [17] CHERNOFF, H. (1959). Sequential design of experiments. *Ann. Math. Statist.* **30** 755–770. [MR0108874](#)
- [18] COVER, T. M. and THOMAS, J. A. (2006). *Elements of Information Theory*, 2nd ed. Wiley, Hoboken, NJ. [MR2239987](#)
- [19] CSISZÁR, I. and SHIELDS, P. C. (2004). Information theory and statistics: A tutorial. *Found. Trends Commun. Inf. Theory* **1** 417–528.
- [20] DEGROOT, M. H. (1962). Uncertainty, information, and sequential experiments. *Ann. Math. Statist.* **33** 404–419. [MR0139242](#)
- [21] DEGROOT, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York. [MR0356303](#)
- [22] DRAGALIN, V. P., TARTAKOVSKY, A. G. and VEERAVALLI, V. V. (1999). Multihypothesis sequential probability ratio tests. I. Asymptotic optimality. *IEEE Trans. Inform. Theory* **45** 2448–2461. [MR1725130](#)
- [23] GALLAGER, R. G. (1968). *Information Theory and Reliable Communication*. Wiley, New York.
- [24] GOEL, P. K. and DEGROOT, M. H. (1979). Comparison of experiments and information measures. *Ann. Statist.* **7** 1066–1077. [MR0536509](#)
- [25] HAROUTUNIAN, E. A., HAROUTUNIAN, M. E. and HARUTYUNYAN, A. N. (2007). Reliability criteria in information theory and in statistical hypothesis testing. *Found. Trends Commun. Inf. Theory* **4** 97–263.
- [26] HAYASHI, M. (2009). Discrimination of two channels by adaptive methods and its application to quantum system. *IEEE Trans. Inform. Theory* **55** 3807–3820. [MR2598075](#)
- [27] HERO, A. O. and COCHRAN, D. (2011). Sensor management: Past, present, and future. *IEEE Sens. J.* **11** 3064–3075.
- [28] HOLLINGER, G. A., MITRA, U. and SUKHATME, G. S. (2011). Active classification: Theory and application to underwater inspection. In *Proceedings of the 15th International Symposium on Robotics Research (ISRR), August 28–September 1, 2011*. Flagstaff, AZ.
- [29] HORSTEIN, M. (1963). Sequential transmission using noiseless feedback. *IEEE Trans. Inform. Theory* **9** 136–143.
- [30] KADANE, J. B. (1971). Optimal whereabouts search. *Oper. Res.* **19** 894–904. [MR0302166](#)
- [31] KEENER, R. (1984). Second order efficiency in the sequential design of experiments. *Ann. Statist.* **12** 510–532. [MR0740909](#)
- [32] KIEFER, J. and SACKS, J. (1963). Asymptotically optimum sequential inference and design. *Ann. Math. Statist.* **34** 705–750. [MR0150907](#)
- [33] KUMAR, P. R. and VARAIYA, P. (1986). *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Prentice-Hall, Upper Saddle River, NJ.
- [34] LALLEY, S. P. and LORDEN, G. (1986). A control problem arising in the sequential design of experiments. *Ann. Probab.* **14** 136–172. [MR0815963](#)
- [35] LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419–1455. [MR0207093](#)
- [36] LEHMANN, E. L. (1988). Comparing location experiments. *Ann. Statist.* **16** 521–533. [MR0947560](#)
- [37] LINDLEY, D. V. (1956). On a measure of the information provided by an experiment. *Ann. Math. Statist.* **27** 986–1005. [MR0083936](#)
- [38] LORDEN, G. (1970). On excess over the boundary. *Ann. Math. Statist.* **41** 520–527. [MR0254981](#)

- [39] LORDEN, G. (1977). Nearly-optimal sequential tests for finitely many parameter values. *Ann. Statist.* **5** 1–21. [MR0438610](#)
- [40] LORDEN, G. (1983). Asymptotic efficiency of three-stage hypothesis tests. *Ann. Statist.* **11** 129–140. [MR0684871](#)
- [41] MCDIARMID, C. (1989). On the method of bounded differences. In *Surveys in Combinatorics, 1989 (Norwich, 1989)*. *London Mathematical Society Lecture Note Series* **141** 148–188. Cambridge Univ. Press, Cambridge. [MR1036755](#)
- [42] NACHLAS, J. A., LONEY, S. R. and BINNEY, B. A. (1990). Diagnostic-strategy selection for series systems. *IEEE Trans. Reliab.* **39** 273–280.
- [43] NAGHSHVAR, M. and JAVIDI, T. (2010). Active M -ary sequential hypothesis testing. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT), 13–18 June 2010* 1623–1627, Austin, TX.
- [44] NAGHSHVAR, M. and JAVIDI, T. (2013). Supplement to “Active sequential hypothesis testing.” DOI:10.1214/13-AOS1144SUPP.
- [45] NAKIBOĞLU, B. and GALLAGER, R. G. (2008). Error exponents for variable-length block codes with feedback and cost constraints. *IEEE Trans. Inform. Theory* **54** 945–963. [MR2445044](#)
- [46] NAKIBOĞLU, B. and ZHENG, L. (2012). Errors-and-erasures decoding for block codes with feedback. *IEEE Trans. Inform. Theory* **58** 24–49. [MR2907699](#)
- [47] NITINAWARAT, S., ATIA, G. and VEERAVALLI, V. V. (2013). Controlled sensing for multihypothesis testing. *IEEE Trans. Automat. Control* **58** 2451–2464.
- [48] NOWAK, R. D. (2011). The geometry of generalized binary search. *IEEE Trans. Inform. Theory* **57** 7893–7906. [MR2895367](#)
- [49] POLYANSKIY, Y., POOR, H. V. and VERDÚ, S. (2011). Feedback in the non-asymptotic regime. *IEEE Trans. Inform. Theory* **57** 4903–4925. [MR2849094](#)
- [50] POLYANSKIY, Y. and VERDU, S. (2011). Hypothesis testing with feedback. In *Information Theory and Applications Workshop (ITA), 6–11 February 2011*, San Diego, CA.
- [51] POSNER, E. (1963). Optimal search procedures. *IEEE Trans. Inform. Theory* **9** 157–160.
- [52] SEJDINOVIC, D. and JOHNSON, O. (2010). Note on noisy group testing: Asymptotic bounds and belief propagation reconstruction. In *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing, September 29–October 1, 2010* 998–1003, Monticello, IL.
- [53] SHANNON, C. E. (1956). The zero error capacity of a noisy channel. *Institute of Radio Engineers Transactions on Information Theory* **2** 8–19. [MR0089131](#)
- [54] SHENOY, P. and YU, A. J. (2011). Rational decision-making in inhibitory control. *Front. Human Neurosci.* **5** 48.
- [55] STONE, L. D. (1975). *Theory of Optimal Search*. Academic Press, New York. [MR0439145](#)
- [56] TOGNETTI, K. P. (1968). An optimal strategy for a whereabouts search. *Oper. Res.* **16** 209–211.
- [57] TORGERSEN, E. (1991). Stochastic orders and comparison of experiments. In *Stochastic Orders and Decision Under Risk (Hamburg, 1989)*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **19** 334–371. IMS, Hayward, CA. [MR1196064](#)
- [58] WALD, A. and WOLFOVITZ, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19** 326–339. [MR0026779](#)

QUALCOMM TECHNOLOGIES, INC.
CORPORATE R&D
5775 MOREHOUSE DRIVE
SAN DIEGO, CALIFORNIA 92121
USA
E-MAIL: mnaghshvar@qti.qualcomm.com

DEPARTMENT OF ELECTRICAL
AND COMPUTER ENGINEERING
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093
USA
E-MAIL: tjavid@ucsd.edu