Sequential Controlled Sensing for Composite Multihypothesis Testing

Aditya Deshmukh Srikrishna Bhashyam Venugopal V. Veeravalli

Abstract

The problem of multi-hypothesis testing with controlled sensing of observations is considered. The distribution of observations collected under each control is assumed to follow a single-parameter exponential family distribution. The goal is to design a policy to find the true hypothesis with minimum expected delay while ensuring that probability of error is below a given constraint. The decision maker can control the delay by intelligently choosing the control for observation collection in each time slot. We derive a policy that satisfies the given constraint on the error probability. We also show that the policy is asymptotically optimal in the sense that it asymptotically achieves an information-theoretic lower bound on the expected delay.

I. Introduction

Sequential controlled sensing is a stochastic framework wherein a decision-maker collects observations from a set of controls by sequentially choosing a control and obtaining an observation associated with that control. This paradigm is encountered in information-gathering systems with multiple degrees of freedom that can be controlled adaptively to achieve a given statistical inference task. In traditional control systems, the control is responsible for governing the state of the system. On the other hand, in controlled sensing, the control governs the quality of observations.

Some applications of controlled sensing are target detection, tracking, classification and dynamic sensor selection. A widely studied problem that can be considered as a special case of controlled sensing is that of anomaly detection. Some applications of anomaly detection include identification of defective batches in manufacturing, detection of abnormal behaviour of machines, and outlier detection in datasets. Another problem studied by the computer science community, which can also be considered to be a special case of controlled sensing, is best arm identification in multi-armed bandits. Controlled sensing has potential applications in diagnostic inference [2], particularly clinical decision support systems, which help clinicians in taking diagnostic decisions. Taking measurements from medical sensors can be expensive and so a potential inference problem would be to find a sequential policy to minimize the number of measurements taken to find the correct hypothesis related to a patient's state of health, with high probability.

We consider the problem of finding the true hypothesis from a finite set of composite hypotheses, with minimum expected delay in a sequential controlled sensing setting, while ensuring that a constraint on the probability of error is satisfied. To achieve this goal, the decision-maker has to intelligently choose a control at each time step in order to make best use of the observations, decide when to stop, and find an appropriate estimate of the true hypothesis.

A. Related Work

Chernoff pioneered controlled sensing in his seminal work [3]. Chernoff considered the problem of composite binary hypothesis testing in a sequential controlled sensing setting. He assumed that the distributions under both hypotheses were parametrized and the two sets of parameters under the hypotheses were disjoint and finite. The set of controls was assumed to be finite as well. Chernoff proposed a policy, known as 'Procedure A', and proved that it is asymptotically optimal under certain positivity constraints on Kullback-Leibler divergences. Albert [4] extended Chernoff's results to the case where the parameter space is infinite with certain restrictions. Bessler [5] also generalized Chernoff's work to multiple hypothesis and an infinite set of controls, but with a finite parameter space. In these papers, the authors named the control sensing problem as 'sequential design of experiments'.

Nitinawarat et al. [6] studied the problem of controlled sensing for multihypothesis testing in a setting where the distributions were assumed known, and provided an asymptotically optimal policy without the positivity assumption of prior work, and with strict guarantees on the probabilities (risks) of choosing the hypotheses incorrectly. Naghshvar et al. [7] considered controlled sensing for sequential multihypothesis in the non-asymptotic regime and analyzed a dynamic programming solution to find the structure of the optimal test, and also studied the problem where the number of hypotheses goes to infinity. The authors of [7] term the controlled sensing problem as 'active sequential hypothesis testing'.

We now discuss related work in anomaly detection, which is a special case of controlled sensing. Li et al. [8] studied outlier hypothesis testing in a setting where there is no control and all processes (taking values in finite sets) are sampled together, and provided a universally exponentially consistent policy when both anomalous and non-anomalous distributions

A part of this work was presented at the 2018 Asilomar Conference on Signals, Systems, and Computers under the title 'Controlled Sensing for Composite Multihypothesis Testing with Application to Anomaly Detection' [1].

are unknown. Cohen et al. [9] considered the problem of anomaly detection with control when both anomalous and non-anomalous distributions are known, and provided an asymptotically optimal deterministic test. Vaidhiyan et al. [10] studied the problem of detecting an odd process among a group of Poisson point processes, in a setting where parameters of the odd and non-odd processes were unknown, and provided an asymptotically optimal policy. Prabhu et al. [11] generalized [10] to vector-exponential families and also considered switching costs.

Best arm identification in multi-armed bandits is a problem well studied by the computer science community. The framework of multi-armed bandits is similar to that of controlled sensing. Kaufmann et al. [12] studied the complexity of identifying best arms in a multi-armed bandit. Garivier et al. [13] provided an asymptotically optimal policy for best arm identification in multi-armed bandits where the distributions on the arms were assumed to belong to a single-parameter exponential family, and the parameters of these distributions were unknown.

B. Paper Outline

In Section II we introduce the problem model. In Section III we provide a lower bound on the expected delay of policies in the class of interest. In Section IV, we give an overview of results and some applications. In Section V, we discuss a proposed policy. In Section VI, we provide some simulations and numerical results. Proofs of all results can be found in Appendices A, B and C.

II. Problem Model

A. Single parameter exponential family

The single parameter exponential family is a collection of probability distributions whose probability density/mass functions can be expressed as

$$p(y;\theta) = h(y)\exp(\theta T(y) - A(\theta)), \tag{1}$$

where θ is the parameter, (also known as the natural parameter) from some parameter set $\Psi \subset \mathbb{R}$, $T : \mathbb{R} \to \mathbb{R}$ represents the statistic, $A : \Theta \to \mathbb{R}$ is a convex function, known as the log-partition function. $A(\theta)$ can be expressed as

$$A(\theta) = \log \int_{-\infty}^{\infty} h(y)\exp(\theta T(y))dy. \tag{2}$$

The distribution can also be parametrized by the expectation parameter κ which is the expected value of the statistic,

$$\kappa = \mathbb{E}_\theta[T(Y)] = \dot{A}(\theta), \tag{3}$$

where $\dot{f}$ is used to represent the derivative of a real-valued function f, that is, $\dot{f} = \frac{df}{dy}$. It is known that A is infinitely differentiable over the domain Ψ.

Let b be the convex conjugate function of A,

$$b(\kappa) = \sup_\theta(\kappa\theta - A(\theta)). \tag{4}$$

Then θ corresponding to κ is given by

$$\theta = \dot{b}(\kappa). \tag{5}$$

The dual relationship between κ and θ is given by

$$\kappa = \dot{A}(\theta) \text{ and } \theta = \dot{b}(\kappa). \tag{6}$$

The KL-divergence between two distributions having natural parameters θ and θ' respectively is given by :

$$D(\theta||\theta') \coloneqq \int_{-\infty}^{\infty} f(y;\theta)\log\frac{f(y;\theta)}{f(y;\theta')}dy \tag{7}$$

$$= A(\theta') - A(\theta) - \dot{A}(\theta)(\theta' - \theta). \tag{8}$$

B. Problem setup

We consider a set of controls denoted by the finite set

$$\mathcal{U} = \{1, 2, 3, \ldots, |\mathcal{U}|\}. \tag{9}$$

The state of nature is denoted by a vector of parameters $\boldsymbol{\theta}$. In the general setting of controlled sensing, the observations under a control, say u, are assumed to follow a non-specific distribution with density, which we denote by $p(y; \boldsymbol{\theta}, u)$, with respect to some common measure μ. In this work, we assume that $\boldsymbol{\theta} = \big(\theta_1, \theta_2, \ldots, \theta_{|\mathcal{U}|}\big)$ is a $|\mathcal{U}|$-dimensional vector and that the distribution of the observations under control u is a member of a single-parameter exponential family with parameter as the u-th coordinate of $\boldsymbol{\theta}$, represented as $\theta_u \in \Psi_u$, where $\Psi_u \subset \mathbb{R}$. Let the domain of $\boldsymbol{\theta}$ be denoted as:

$$\Omega = \Psi_1 \times \Psi_2 \times \ldots \Psi_{|\mathcal{U}|}. \tag{10}$$

The probability density/mass function of observation y under control u and given parameters $\boldsymbol{\theta}$ is given by

$$p(y; \boldsymbol{\theta}, u) = h_u(y) \exp[\theta_u T_u(y) - A_u(\theta_u)], \tag{11}$$

where T_u is the statistic function and A_u is the log-partition function of the exponential family associated with control u. Let b_u be the convex conjugate function of A_u. The KL-divergence between between the distributions under control u and u', for control parameters $\boldsymbol{\theta}$ and $\boldsymbol{\theta'}$ is

$$D_u(\boldsymbol{\theta} || \boldsymbol{\theta'}) \coloneqq \int_{-\infty}^{\infty} p(y; \boldsymbol{\theta}, u) \log \frac{p(y; \boldsymbol{\theta}, u)}{p(y; \boldsymbol{\theta'}, u)} d\mu(y) \tag{12}$$

$$= A_u(\theta'_u) - A_u(\theta_u) - \dot{A}_u(\theta_u)(\theta'_u - \theta_u). \tag{13}$$

The set of hypothesis is denoted by

$$\mathcal{M} = \{1, 2, 3, \ldots, M\}. \tag{14}$$

Under hypothesis $m \in \mathcal{M}$, $\boldsymbol{\theta} \in \Theta_m$, where $\Theta_m \subset \Omega$. Let $\|.\|$ be a norm on $\mathbb{R}^{|\mathcal{U}|}$. We assume the following structure on the sets Θ_m.

1) Each Θ_m is a disjoint finite union of sets, that is $\Theta_m = \bigcup_{i=1}^{x_m} \Gamma_m^{(i)}$, where $x_m \in \mathbb{N}$ and $\Gamma_m^{(i)} \cap \Gamma_m^{(j)} = \phi$, $\forall i, j \in [x_m]$ such that $i \neq j$.
2) $\forall m \in \mathcal{M}, \forall i \in [x_m], \Gamma_m^{(i)}$ is convex and open in its own affine hull, denoted by $\text{aff}\Big(\Gamma_m^{(i)}\Big)$.
3) $\forall m_1, m_2 \in \mathcal{M}$ such that $m_1 \neq m_2$, we have $\Gamma_{m_1}^{(i)} \cap \Gamma_{m_2}^{(j)} = \phi$, $\forall i \in [x_{m_1}]$, $\forall j \in [x_{m_2}]$. Note that this implies Θ_m's are mutually disjoint.
4) $\forall m \in \mathcal{M}, \forall i \in [x_m], \forall \boldsymbol{\theta} \in \Gamma_m^{(i)}$, $\Delta\Big(\boldsymbol{\theta}, \Gamma_{m'}^{(j)}\Big) > 0$ for any $m' \in \mathcal{M}, j \in [x_{m'}]$ such that $m' \neq m$ or $j \neq i$, where $\Delta(\boldsymbol{x}, A) : \Omega \to \mathbb{R}$ is the distance of $\boldsymbol{x} \in \Omega$ to the set $A \subset \Omega$ given by

$$\Delta(\boldsymbol{x}, A) \coloneqq \inf\{\|\boldsymbol{x} - \boldsymbol{\theta'}\| : \boldsymbol{\theta'} \in A\}. \tag{15}$$

We consider a sequential setting where at each time step $k = 1, 2, 3, \ldots$ the controller selects a control U_k and gets an observation Y_k. All observations $(Y_k)_{k \geq 1}$ and all control selections $(U_k)_{k \geq 1}$ are assumed to be defined on a common probability space. Let $\mathcal{F}_n = \sigma(U_1, Y_1, U_2, Y_2, \ldots, U_n, Y_n)$ be the sigma-algebra generated by the selected controls and observations up to time n. $\mathbb{P}_{\boldsymbol{\theta}}[.]$ and $\mathbb{E}_{\boldsymbol{\theta}}[.]$ denote the probability and expectation respectively conditioned that the vector of parameters is $\boldsymbol{\theta}$. A policy $\Phi = (\{U_n\}, \tau, \hat{m})$ is then defined by:

- a sequence of controls $\{U_n\}$, where U_n is $\mathcal{F}_{n-1}$ measurable,
- a stopping rule τ, which is a stopping time with respect to $U_1, Y_1, U_2, Y_2, \ldots$, and
- an $\mathcal{F}_\tau$-measurable decision $\hat{m}$ which is the policy's estimate of the true hypothesis.

Any such policy keeps taking observations by choosing controls based on past observations and chosen controls, until the stopping time. At the stopping time, the policy stops taking any further observations and choosing any further controls, and outputs an estimate of the true hypothesis. The goal is to design a policy to find the true hypothesis with minimum expected delay $\mathbb{E}_{\boldsymbol{\theta}}[\tau]$ while ensuring that probability of error is below a given constraint α. Let

$$g(\boldsymbol{\theta}) \coloneqq m \text{ if } \boldsymbol{\theta} \in \Theta_m. \tag{16}$$

Definition 1 ($\bar{\alpha}$-correct policy)**.** Let $\alpha \in (0, 1)$. A policy is called $\bar{\alpha}$-correct if $\forall \boldsymbol{\theta} \in \bigcup_{m=1}^{M} \Theta_m$, $\mathbb{P}_{\boldsymbol{\theta}}[\tau < \infty] = 1$ and $\mathbb{P}_{\boldsymbol{\theta}}[\hat{m} \neq g(\boldsymbol{\theta})] \leq \alpha$.

For any state of nature parameters, an $\bar{\alpha}$-correct policy stops in finite time almost surely and detects the true hypothesis with probability of at-least $1-\alpha$. We contribute a policy which we show to be $\bar{\alpha}$-correct and asymptotically optimal in the sense that it achieves the aymptotic lower bound on expected delay as $\alpha \to 0$. We discuss this lower bound in the next section.

III. Lower bound

We first establish a lower bound on the expected delay of any $\bar{\alpha}$-correct policy.

Lemma 1. Let $\alpha \in (0,1)$. Then $\forall \boldsymbol{\theta} \in \bigcup_{m=1}^{M} \Theta_m$, any $\bar{\alpha}$-correct policy satisfies

$$\mathbb{E}_{\boldsymbol{\theta}}[\tau] \geq \frac{d(\alpha||1-\alpha)}{D^*(\boldsymbol{\theta})}, \tag{17}$$

where $D^*(\boldsymbol{\theta})$ is defined as,

$$D^*(\boldsymbol{\theta}) \coloneqq \sup_{\boldsymbol{q}\in\mathcal{P}} \inf_{\boldsymbol{\theta}' \in \bigcup_{m=1}^{M} \Theta_m \setminus \Theta_{g(\boldsymbol{\theta})}} \sum_{u=1}^{|\mathcal{U}|} q_u D_u(\boldsymbol{\theta}||\boldsymbol{\theta}'). \tag{18}$$

$d(x||y) \coloneqq x\log\left(\frac{x}{y}\right) + (1-x)\log\left(\frac{1-x}{1-y}\right)$ represents the binary relative entropy function and the supremum is taken over $\mathcal{P}$, the set of all distributions over $\mathcal{U}$.

Proof. The proof follows from Lemma 1 in [12], which is stated for multi-armed bandit models, but can be applied to the case of sequential controlled sensing due to similarity in the paradigms. □

We further analyze $D^*(\boldsymbol{\theta})$ to gain insights and discover properties which might help us in designing a good policy for the problem in consideration.

Proposition 1. The supremum in (18) is a maximum and attained $\forall \boldsymbol{\theta} \in \bigcup_{m=1}^{M} \Theta_m$ at

$$\boldsymbol{q}^*(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{q}\in\mathcal{P}} \inf_{\boldsymbol{\theta}' \in \bigcup_{m=1}^{M} \Theta_m \setminus \Theta_{g(\boldsymbol{\theta})}} \sum_{u=1}^{|\mathcal{U}|} q_u D_u(\boldsymbol{\theta}||\boldsymbol{\theta}'). \tag{19}$$

Furthermore, $\boldsymbol{q}^*(\boldsymbol{\theta})$ is continuous at each $\boldsymbol{\theta}$.

Proof. See Appendix A for the proof. We assume that the hypothesis sets are such that the maximum is unique. In the case of best arm identification [Theorem 5, [13]] and anomaly detection [Proposition 3, [11]], the maximum is indeed unique. □

Some remarks are in order: First, the lower bound in (17) is non-asymptotic in nature, and so it is a stronger result than the asymptotic lower bounds generally seen in the literature on controlled sensing and anomaly detection, see, e.g., [3], [6] and [9]. Taking the limit as $\alpha \to 0$, the asymptotic lower bound we get is

$$\liminf_{\alpha\to 0} \frac{\mathbb{E}_{\boldsymbol{\theta}}[\tau]}{|\log\alpha|} \geq \frac{1}{D^*(\boldsymbol{\theta})}, \tag{20}$$

which has the same form as the asypmtotic lower bounds generally found in controlled sensing literature. Moreover, this lower bound is applicable not just to single-parameter exponential distributions, but to general parametrized families. Intituively, $q_u^*(\boldsymbol{\theta})$ represents the optimal proportion of the number of times control u should be chosen by a policy that tries to achieve the lower bound, and $D^*(\boldsymbol{\theta})$ represents the maximum possible rate of 'information' extraction in the worst case scenario.

IV. Overview of results

We propose a policy, based on the policy given in [13], and show the following properties.

Theorem 1. The proposed policy is an $\bar{\alpha}$-correct policy for any given $\alpha \in (0,1)$.

Theorem 2. [Almost-sure upper bound] For any state of nature $\boldsymbol{\theta} \in \bigcup_{m=1}^{M} \Theta_m$, the proposed policy satisfies

$$\mathbb{P}_{\boldsymbol{\theta}}\left[\limsup_{\alpha\to 0} \frac{\tau}{|\log\alpha|} \leq \frac{1}{D^*(\boldsymbol{\theta})}\right] = 1. \tag{21}$$

Theorem 3. [Asymptotic optimality in expectation] For any state of nature $\boldsymbol{\theta} \in \bigcup_{m=1}^{M} \Theta_m$, the proposed policy satisfies

$$\limsup_{\alpha \to 0} \frac{\mathbb{E}_{\boldsymbol{\theta}}[\tau]}{|\log \alpha|} \leq \frac{1}{D^*(\boldsymbol{\theta})}. \tag{22}$$

Proofs of the above theorems are given in Appendix B. Theorem 3 implies that the proposed policy is asymptotically optimal. We now discuss two applications of composite multihypothesis controlled sensing. In both applications we assume that the distributions of observations collected across all controls follow the same single-parameter exponential family.

1) Best-K arms identification in a multi-armed bandit: The problem of identification of best-K arms in a multi-armed bandit with minimum expected delay under constraint on error probability, can be cast as a sequential controlled sensing problem where each control corresponds to an arm and there are $M = \binom{|\mathcal{U}|}{K}$ hypotheses, such that each hypothesis consists of a unique combination of K arms which have the highest expectation parameters (we assume that the statistic T is identity as in [13]). Since $\dot{A}$ is increasing, we can express any hypothesis set as

$$\Theta = \bigcup_{\pi} \Big\{ \boldsymbol{\theta}' \in \Omega : \theta'_{\pi(\phi(1))} \geq \theta'_{\pi(\phi(2))} \geq \cdots \geq \theta'_{\pi(\phi(K))}, \forall i \notin \{\phi(1), \phi 2, \ldots, \phi(K)\} \Big\} \tag{23}$$

where $\{\phi(1), \phi 2, \ldots, \phi(K)\}$ denotes a unique combination of K controls and π denotes a permutation of this combination. Note that each hypothesis set is open and convex. All hypothesis sets are mutually disjoint. Hence, Assumptions 1, 2 and 3 hold. It can be verified that Assumption 4 also holds. Thus the proposed policy can be applied in this scenario and we get an asymptotically optimal policy.

2) Sequential controlled anomaly detection: The framework of controlled anomaly detection consists of multiple streams of observations. All distributions are the same except for one stream, which we call as the anomalous stream. The objective is to sequentially collect observations by choosing one stream in each time step, and find the anomalous stream in minimum expected delay, while ensuring that the probability of error is bounded by a given constraint. We can cast the anomaly detection problem as a controlled sensing problem where each control picks a unique stream to collect observations, and the hypotheses are as follows. Let $M = |\mathcal{U}|$ and for $m \in \mathcal{M}$,

$$\Theta_m = \{\boldsymbol{\theta}' \in \Omega : \theta'_i = \theta, \forall i \neq m \text{ for some } \theta \text{ and } \theta'_m \neq \theta\}. \tag{24}$$

So, hypothesis m indicates that the m^{th} stream is anomalous. Observe that each Θ_m is a 2-D plane (the degrees of freedom being the anomalous and non-anomalous parameters) without the 1-D line given by

$$L = \Big\{ \boldsymbol{\theta}' \in \Omega : \theta'_1 = \theta'_2 = \cdots = \theta'_{|\mathcal{U}|} \Big\}. \tag{25}$$

Thus each Θ_m can be expressed as a union of two convex sets which are open in their own affine hulls, and all such convex sets that form the hypothesis sets are mutually disjoint. Hence, Assumptions 1, 2 and 3 hold. It can be verified that Assumption 4 also holds. Thus the proposed policy can be applied in this scenario and we get an asymptotically optimal policy.

V. Proposed policy

Recall that a policy has three essential components: a decision, a control law and a stopping rule. We discuss these components in detail, after introducing the required notation.

Let $N_u(n)$ be the number of times control u is chosen up to time n.

$$N_u(n) = \sum_{k=1}^{n} \mathbb{1}_{\{U_k = u\}}. \tag{26}$$

Let $S_u(n)$ be the sum of sufficient statistics of control u up to time n.

$$S_u(n) = \sum_{k=1}^{n} T_u(Y_k) \mathbb{1}_{\{U_k = u\}}. \tag{27}$$

For all hypothesis $i, j \in \mathcal{M}$, we define the Generalized Likelihood Ratio Test Statistic as

$$Z_{i,j}(n) := \log \frac{\sup_{\boldsymbol{\theta}' \in \Theta_i} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)}{\sup_{\boldsymbol{\theta}'' \in \Theta_j} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}'', u)}, \tag{28}$$

where $\underline{Y}^u(n) = (Y_k : U_k = u, k \leq n)$ is the collection of observations from control u. Let

$$Z_i(n) \coloneqq \min_{j \neq i, j \in \mathcal{M}} Z_{i,j}(n) \text{ and } Z(n) \coloneqq \max_{i \in \mathcal{M}} Z_i(n). \tag{29}$$

Let $\boldsymbol{\theta}^*(n) \in \Omega$ be the global maximum likelihood estimate of $\boldsymbol{\theta}$,

$$\boldsymbol{\theta}^*(n) \coloneqq \arg\max_{\boldsymbol{\theta}' \in \Omega} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u). \tag{30}$$

So, $\forall u \in \mathcal{U}$,

$$\theta_u^*(n) = b_u\left(\frac{S_u(n)}{N_u(n)}\right). \tag{31}$$

A. Stopping time

We adopt the approach in [13] and define the stopping time as follows.

$$\tau \coloneqq \inf\{n \in \mathbb{N} : Z(n) \geq \beta(n, \alpha)\}. \tag{32}$$

$\beta(n, \alpha)$ is a dynamic threshold given by

$$\beta(n, \alpha) = v(n) + w(\alpha), \tag{33}$$

where

$$w(\alpha) = |\log \alpha| + \sqrt{4|\mathcal{U}||\log \alpha|} \tag{34}$$

and

$$v(n) = C + \log\Big(n(1 + \log n)^{|\mathcal{U}|+2}\Big) + \sqrt{4|\mathcal{U}| \log\big(n(1 + \log n)^{|\mathcal{U}|+2}\big)}. \tag{35}$$

Here C is a constant given by $C = 2|\mathcal{U}|\sqrt{2\log\frac{2|\mathcal{U}|}{e} + \frac{1}{|\mathcal{U}|}\log\frac{2e^{|\mathcal{U}|+1}}{|\mathcal{U}|^{|\mathcal{U}|}}} + \log\frac{2e^{|\mathcal{U}|+1}}{|\mathcal{U}|^{|\mathcal{U}|}}$.

The threshold $\beta(n, \alpha)$ is based on the deviation inequality given in Theorem 2 in [14], which is stated for Bernoulli distributions, but easily extendable to single-parameter exponential family distributions. For sake of completeness, we provide this extension in Appendix C.

B. Decision

At each time step, the policy's estimate of the true hypothesis will be called as the recommendation at that time step. The recommendation $\hat{r}(n)$ is the nearest hypothesis set Θ_m to the global MLE $\boldsymbol{\theta}^*(n)$.

$$\hat{r}(n) \in \arg\min_{m \in \mathcal{M}} \Delta(\boldsymbol{\theta}^*(n), \Theta_m). \tag{36}$$

The decision is given by:

$$\hat{m} \in \arg\max_{m \in \mathcal{M}} Z_m(\tau). \tag{37}$$

C. Control Law

For initialization, all controls are selected once. For the control law, we follow the approach used in the 'track-and-stop' strategy, proposed in [13]. The idea is to choose the control so as to get the empirical proportions $\left(\frac{N_u(n)}{n}\right)$ close to the optimal proportions $\boldsymbol{q}^*(\boldsymbol{\theta})$. Since $\boldsymbol{\theta}$ is unknown, we use the plug-in estimates $\boldsymbol{q}^*(\hat{\boldsymbol{\theta}}(n))$, where $\hat{\boldsymbol{\theta}}(n)$ is the nearest vector in recommended hypothesis set $\Theta_{\hat{r}(n)}$ to the global MLE $\boldsymbol{\theta}^*(n)$.

$$\hat{\boldsymbol{\theta}}(n) \in \arg\min_{\boldsymbol{\theta}' \in \Theta_{\hat{r}(n)}} \|\boldsymbol{\theta}' - \boldsymbol{\theta}^*(n)\|. \tag{38}$$

If no minimizer exists, choose $\hat{\boldsymbol{\theta}}(n)$ to be ρ-closest of $\boldsymbol{\theta}^*(n)$ in $\Theta_{\hat{r}(n)}$, where $\rho > 1$ is fixed.

$$\left\|\hat{\boldsymbol{\theta}}(n) - \boldsymbol{\theta}^*(n)\right\| \leq \rho \inf_{\boldsymbol{\theta}' \in \Theta_{\hat{r}(n)}} \|\boldsymbol{\theta}' - \boldsymbol{\theta}^*(n)\| = \rho\Delta(\boldsymbol{\theta}^*(n), \Theta_{\hat{r}(n)}). \tag{39}$$

Let $\boldsymbol{q}^\epsilon(\boldsymbol{\theta})$ be a L^∞ projection of $q^*(\boldsymbol{\theta})$ onto $\{\boldsymbol{q} \in \mathcal{P} : \forall u \in \mathcal{U}, q_u \in [\epsilon, 1]\}$. Then we select the control at time $n+1$ according to

$$u_{n+1} \in \arg\max_{u \in \mathcal{U}} \sum_{k=1}^{n} q_u^{\epsilon_k}(\hat{\boldsymbol{\theta}}(k)) - N_u(n), \tag{40}$$

where $\epsilon_k = \frac{1}{2}(|\mathcal{U}|^2 + k)^{-1/2}$. Note that this projection enforces exploration of the controls in the initial stages when the estimates are not quite accurate. This forced exploration decays as time progresses.

Lemma 2 (Lemma 7, [13])**.** The control law ensures that $\forall n \in \mathbb{N}$ and $\forall u \in \mathcal{U}$,

$$N_u(n) \geq \sqrt{n + |\mathcal{U}|^2} - 2|\mathcal{U}| \tag{41}$$

and that

$$\max_{u \in \mathcal{U}} \left| N_u(n) - \sum_{k=0}^{n-1} q_u^*(\hat{\boldsymbol{\theta}}(k)) \right| \leq |\mathcal{U}|(1 + \sqrt{n}). \tag{42}$$

Observe that the GLRT statistic has a maximum likelihood in the numerator, which makes it difficult to find a constant threshold such that probability of error can be constrained. In [10], for example, the authors circumvent this problem by defining a modified GLRT statistic, which has a likelihood averaged over a prior in the numerator instead of the maximum likelihood, and have a constant threshold policy.

VI. Numerical results

We implemented the proposed policy in a general composite multi-hypothesis detection scenario. The set of controls is $\mathcal{U} = \{1, 2, 3, 4, 5\}$. The observations from the controls follow normal distributions with means $\boldsymbol{\mu} = \{1, 2, 12, 8, 15\}$ and variances $\boldsymbol{\sigma}^2 = \{1, 1, 16, 4, 9\}$. The variances are assumed to be known. In this case, the true parameter for control u is $\theta_u = \frac{\mu_u}{\sigma_u}$. So, the true vector of parameters is $\boldsymbol{\theta} = \{1, 2, 3, 4, 5\}$. The hypothesis are as follows:

$$\Theta_1 = \{0 \leq \theta_1 \leq 2, 1 \leq \theta_2 \leq 3, 2 \leq \theta_3 \leq 4, 3 \leq \theta_4 \leq 5, 4 \leq \theta_5 \leq 6\} \tag{43}$$

$$\Theta_2 = \{0 \leq \theta_1 \leq 2, -2 \leq \theta_2 \leq 0, 4 \leq \theta_3 \leq 6, 3 \leq \theta_4 \leq 5, 7 \leq \theta_5 \leq 9\} \tag{44}$$

$$\Theta_3 = \{-2 \leq \theta_1 \leq 0, 1 \leq \theta_2 \leq 3, 2 \leq \theta_3 \leq 4, 5 \leq \theta_4 \leq 7, 2 \leq \theta_5 \leq 5\} \tag{45}$$

$$\Theta_4 = \{-2 \leq \theta_1 \leq 0, 3 \leq \theta_2 \leq 5, 0 \leq \theta_3 \leq 2, 3 \leq \theta_4 \leq 5, 4 \leq \theta_5 \leq 6\}. \tag{46}$$

Fig. 1 shows the plot of the ratio of empirical mean stopping time to $|\log(\alpha)|$ versus $|\log(\alpha)|$, in comparison with the lower bound $\frac{1}{D^*(\boldsymbol{\theta})} = 2.2601$. The empirical mean stopping time is the average of the stopping times obtained in 100 independent iterations. Observe that the ratio of the empirical mean stopping time to $|\log(\alpha)|$ approaches the lower bound $\frac{1}{D^*(\boldsymbol{\theta})}$, as α decreases, thereby demonstrating the asymptotic optimality of the proposed policy.

Appendix A

Proof of proposition 1. Let $\boldsymbol{\theta} \in \bigcup_{m=1}^{M} \Theta_m$ be fixed, and $\Theta = \bigcup_{m=1}^{M} \Theta_m \setminus \Theta_{g(\boldsymbol{\theta})}$. Let $f(\boldsymbol{q}) : \mathcal{P} \to \mathbb{R}$, such that

$$f(\boldsymbol{q}) = \inf_{\boldsymbol{\theta}' \in \Theta} \sum_{u=1}^{|\mathcal{U}|} q_u D_u(\boldsymbol{\theta} || \boldsymbol{\theta}'). \tag{47}$$

Note that the map $(\boldsymbol{q}, \boldsymbol{\theta}') \mapsto \sum_{u=1}^{U} q_u D_u(\boldsymbol{\theta} || \boldsymbol{\theta}')$ is bounded below by 0, and Θ is non-empty, so $f(\boldsymbol{q})$ is well-defined. Let

$$\mathcal{P}_S \coloneqq \{\boldsymbol{q} \in \mathcal{P} : \forall u \in S, q_u > 0 \text{ and } \forall u \notin S, q_u = 0\}, \tag{48}$$

for $S \in 2^{\mathcal{U}} \setminus \{\phi\}$. Let $\mathcal{S} = 2^{\mathcal{U}} \setminus \left\{ \{\phi\} \bigcup_{u=1}^{|\mathcal{U}|} \{u\} \right\}$. Note that

$$\mathcal{P} = \bigcup_{S \in 2^{\mathcal{U}} \setminus \{\phi\}} \mathcal{P}_S = \bigcup_{S \in \mathcal{S}} \mathcal{P}_S \bigcup_{u=1}^{|\mathcal{U}|} \mathcal{P}_{\{u\}}. \tag{49}$$

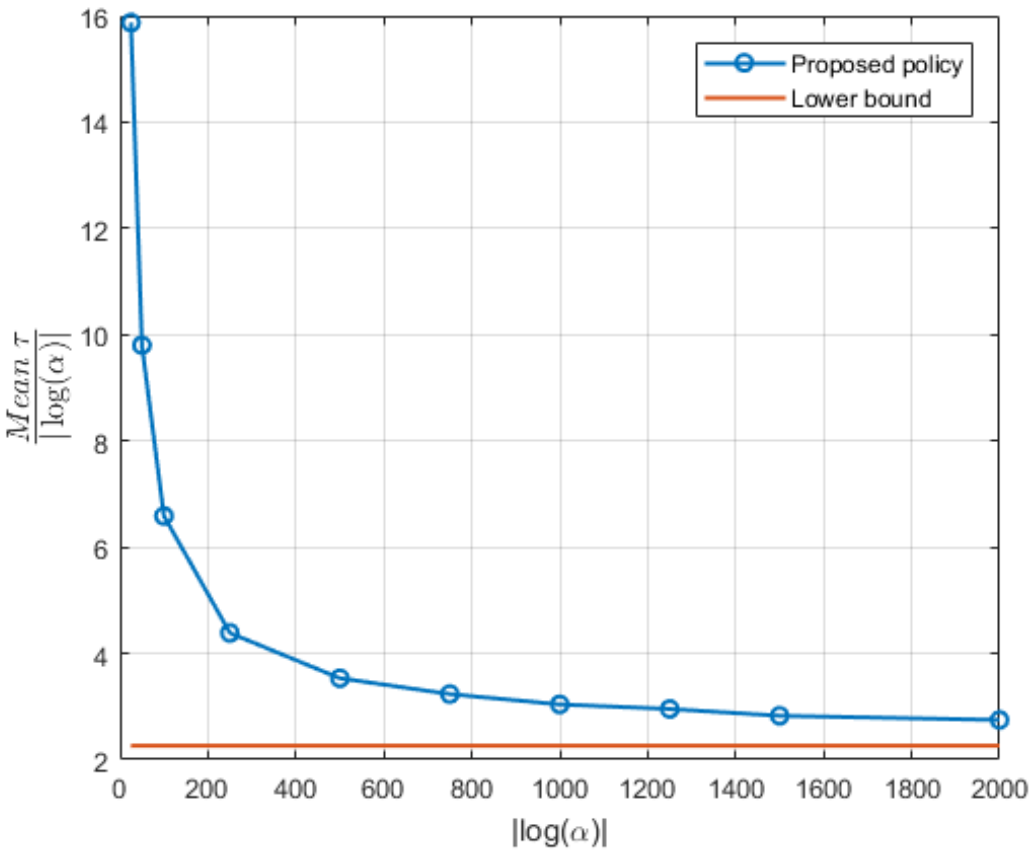

Fig. 1. Performance of proposed policy in a general composite hypothesis setting

Note that f is concave on $\mathcal{P}$, since it is an infima of an affine family of functions. Hence, we have that f is concave on $\mathcal{P}$. Since for any $S \in \mathcal{S}$, $\mathcal{P}_S \subset \mathcal{P}$, f is concave on $\mathcal{P}_S$ and thus continuous on relint$(\mathcal{P}_S) = \mathcal{P}_S$. We now show that f is lower semi-continuous on $\mathcal{P}$. Consider any $S \in \mathcal{S}$. Let $\boldsymbol{q}_0 \in \mathcal{P}_S$. Consider a sequence $\{\boldsymbol{q}_n\} \subset \mathcal{P}$ such that $\boldsymbol{q}_n \to \boldsymbol{q}_0$. So we have,

$$\liminf_{n\to\infty} f(\boldsymbol{q}_n) = \liminf_{n\to\infty} \inf_{\boldsymbol{\theta}'\in\Theta} \left\{ \sum_{u\in S} q_{n,u} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') + \sum_{u\notin S} q_{n,u} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') \right\} \tag{50}$$

$$\geq \liminf_{n\to\infty} \left\{ \inf_{\boldsymbol{\theta}'\in\Theta} \sum_{u\in S} q_{n,u} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') + \inf_{\boldsymbol{\theta}'\in\Theta} \sum_{u\notin S} q_{n,u} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') \right\} \tag{51}$$

$$\geq \liminf_{n\to\infty} \left\{ \left(\sum_{u\in S} q_{n,u} \right) \inf_{\boldsymbol{\theta}'\in\Theta} \sum_{u\in S} \frac{q_{n,u}}{\left(\sum_{u\in S} q_{n,u} \right)} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') + \sum_{u\notin S} q_{n,u} \inf_{\boldsymbol{\theta}'\in\Theta} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') \right\}. \tag{52}$$

Since $\forall u \notin S, q_{n,u} \to q_{0,u} = 0$, we consequently get

$$\sum_{u \notin S} q_{n,u} \inf_{\boldsymbol{\theta}' \in \Theta} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') \to 0, \tag{53}$$

$$\sum_{u \in S} q_{n,u} \to 1, \tag{54}$$

$$\forall u \in S, \frac{q_{n,u}}{\left(\sum_{u \in S} q_{n,u}\right)} \to q_{0,u}, \tag{55}$$

$$\inf_{\boldsymbol{\theta}' \in \Theta} \sum_{u \in S} \frac{q_{n,u}}{\left(\sum_{u \in S} q_{n,u}\right)} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') \to f(\boldsymbol{q}_0). \tag{56}$$

Note that (56) follows from the continuity of f on $\mathcal{P}_S$. Applying (53) and (56) in (52), we get

$$\liminf_{n \to \infty} f(\boldsymbol{q}_n) \geq f(\boldsymbol{q}_0). \tag{57}$$

Now consider the singleton element $\boldsymbol{q}_0 \in \mathcal{P}_{\{v\}}$ for any $v \in \mathcal{U}$. Note that $q_{0,u} = 1$. Similarly as before, consider a sequence $\{\boldsymbol{q}_n\} \subset \mathcal{P}$ such that $\boldsymbol{q}_n \to \boldsymbol{q}_0$. So we have,

$$\liminf_{n \to \infty} f(\boldsymbol{q}_n) = \liminf_{n \to \infty} \inf_{\boldsymbol{\theta}' \in \Theta} \sum_{u=1}^{|\mathcal{U}|} q_{n,u} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') \tag{58}$$

$$\geq \liminf_{n \to \infty} \sum_{u=1}^{|\mathcal{U}|} q_{n,u} \inf_{\boldsymbol{\theta}' \in \Theta} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') \tag{59}$$

$$= \inf_{\boldsymbol{\theta}' \in \Theta} D_v(\boldsymbol{\theta}||\boldsymbol{\theta}') \tag{60}$$

$$= f(\boldsymbol{q}_0). \tag{61}$$

Since f is lower semi-continuous on $\mathcal{P}_S$ for any $S \in \mathcal{S}$ and on $\mathcal{P}_{\{u\}}$ for any $u \in \mathcal{U}$, f is lower semi-continuous on $\mathcal{P}$. We now show that f is upper semi-continuous on $\mathcal{P}$. Consider a sequence $\{\boldsymbol{q}_n\} \subset \mathcal{P}$ such that $\boldsymbol{q}_n \to \boldsymbol{q}_0 \in \mathcal{P}$. From the definition of f, it follows that $\exists \{\boldsymbol{\theta}_k\} \subset \Theta$ such that

$$\sum_{u=1}^{|\mathcal{U}|} q_{0,u} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}_k) \to f(\boldsymbol{q}_0). \tag{62}$$

So we get,

$$\limsup_{n \to \infty} f(\boldsymbol{q}_n) = \limsup_{n \to \infty} \inf_{\boldsymbol{\theta}' \in \Theta} \sum_{u=1}^{|\mathcal{U}|} q_{n,u} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') \tag{63}$$

$$\leq \limsup_{n \to \infty} \sum_{u=1}^{|\mathcal{U}|} q_{n,u} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}_k) \tag{64}$$

$$= \sum_{u=1}^{|\mathcal{U}|} q_{0,u} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}_k). \tag{65}$$

Note that this holds for all $k \in \mathbb{N}$. Thus, taking limit $k \to \infty$, we get

$$\limsup_{n \to \infty} f(\boldsymbol{q}_n) \leq f(\boldsymbol{q}_0). \tag{66}$$

Hence, f is upper semi-continuous on $\mathcal{P}$. Since f is both upper and lower semi-continuous on $\mathcal{P}$, we conclude f is continuous on $\mathcal{P}$. Since $\mathcal{P}$ is compact, f achieves the maximum value $D^*(\boldsymbol{\theta})$ on $\mathcal{P}$.

We now show that the function $f_0(\boldsymbol{q}, \boldsymbol{\theta}) : \mathcal{P} \times \bigcup_{m=1}^{M} \Theta_m \to \mathbb{R}$ given by

$$f_0(\boldsymbol{q}, \boldsymbol{\theta}) = \inf_{\boldsymbol{\theta}' \in \bigcup_{m=1}^{M} \Theta_m \setminus \Theta_{g(\boldsymbol{\theta})}} \sum_{u=1}^{|\mathcal{U}|} q_u D_u(\boldsymbol{\theta}||\boldsymbol{\theta}'), \tag{67}$$

is continuous. Let $m' \in \mathcal{M}$, $i \in \{1, 2, ..., x_{m'}\}$, and $\boldsymbol{q} \in \mathcal{P}$ be fixed. First, we show that the functions $f_j^{(m)} : \Gamma_{m'}^{(i)} \to \mathbb{R}$ given by

$$f_j^{(m)}(\boldsymbol{\theta}) = \inf_{\boldsymbol{\theta}' \in \Gamma_m^{(j)}} \sum_{u=1}^{|\mathcal{U}|} q_u D_u(\boldsymbol{\theta} || \boldsymbol{\theta}') \tag{68}$$

are continuous, where $m \in \mathcal{M}$ and $j \in \{1, 2, \ldots, x_m\}$. We show that $f_j^{(m)}$ is convex as follows. Let $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Gamma_{m'}^{(i)}$ and $\lambda \in [0, 1]$. So for any $\boldsymbol{\theta}_1', \boldsymbol{\theta}_2' \in \Gamma_m^{(j)}$,

$$f_j^{(m)}(\lambda\boldsymbol{\theta}_1 + (1-\lambda)\boldsymbol{\theta}_2) \leq \sum_{u=1}^{|\mathcal{U}|} q_u D_u(\lambda\boldsymbol{\theta}_1 + (1-\lambda)\boldsymbol{\theta}_2 || \lambda\boldsymbol{\theta}_1' + (1-\lambda)\boldsymbol{\theta}_2') \tag{69}$$

$$\leq \sum_{u=1}^{|\mathcal{U}|} q_u [\lambda D_u(\boldsymbol{\theta}_1 || \boldsymbol{\theta}_1') + (1-\lambda) D_u(\boldsymbol{\theta}_2 || \boldsymbol{\theta}_2')] \tag{70}$$

$$= \lambda \sum_{u=1}^{|\mathcal{U}|} q_u D_u(\boldsymbol{\theta}_1 || \boldsymbol{\theta}_1') + (1-\lambda) \sum_{u=1}^{|\mathcal{U}|} q_u D_u(\boldsymbol{\theta}_2 || \boldsymbol{\theta}_2') \tag{71}$$

This holds due to the convexity of D_u. Taking infimum over $\boldsymbol{\theta}_1'$ and $\boldsymbol{\theta}_2'$, we get

$$f_j^{(m)}(\lambda\boldsymbol{\theta}_1 + (1-\lambda)\boldsymbol{\theta}_2) \leq \lambda f_j^{(m)}(\boldsymbol{\theta}_1) + (1-\lambda) f_j^{(m)}(\boldsymbol{\theta}_2). \tag{72}$$

Thus, $f_j^{(m)}$ is convex on $\Gamma_{m'}^{(i)}$ and hence continuous on $\Gamma_{m'}^{(i)}$, since $\Gamma_{m'}^{(i)} = \text{relint}\left(\Gamma_{m'}^{(i)}\right)$. This further implies that $\min_{m \neq m'} \min_{j \in [x_m]} f_j^{(m)}$ is continuous on $\Gamma_{m'}^{(i)}$. Since $m' \in \mathcal{M}$, $i \in [x_{m'}]$ were chosen arbitrarily and from assumption 4 on the structure of Θ_m's, we get that for a fixed $\boldsymbol{q} \in \mathcal{P}$, the function $f^* : \bigcup_{m=1}^{M} \Theta_m \to \mathbb{R}$ given by

$$f^*(\boldsymbol{\theta}) = \min_{m \neq g(\boldsymbol{\theta})} \min_{j \in [x_m]} \inf_{\boldsymbol{\theta}' \in \Gamma_m^{(j)}} \sum_{u=1}^{|\mathcal{U}|} q_u D_u(\boldsymbol{\theta} || \boldsymbol{\theta}') = \inf_{\boldsymbol{\theta}' \in \bigcup_{m=1}^{M} \Theta_m \setminus \Theta_{g(\boldsymbol{\theta})}} \sum_{u=1}^{|\mathcal{U}|} q_u D_u(\boldsymbol{\theta} || \boldsymbol{\theta}') \tag{73}$$

is continuous on its domain. We prove the continuity of $f_0(\boldsymbol{q}, \boldsymbol{\theta})$ by using the continuity of $f(\boldsymbol{q})$ and $f^*(\boldsymbol{\theta})$. Consider a sequence $\{\boldsymbol{q}_n\} \subset \mathcal{P}$ such that $\boldsymbol{q}_n \to \boldsymbol{q}_0 \in \mathcal{P}$ and a sequence $\{\boldsymbol{\theta}_n\} \subset \bigcup_{m=1}^{M} \Theta_m$ such that $\boldsymbol{\theta}_n \to \boldsymbol{\theta}_0 \in \bigcup_{m=1}^{M} \Theta_m$. Let $\boldsymbol{q}_0 \in \mathcal{P}_S$. Note that $\exists N_1$ such that $\forall n \geq N_1$, $\boldsymbol{\theta}_n \in \Theta_{g(\boldsymbol{\theta}_0)}$. Given $\epsilon \in \left(0, \min_{u \in S} \frac{q_{0,u}}{2}\right)$, $\exists N_2$ such that $\forall u \in \mathcal{U}, \forall n \geq N_2$, $q_{n,u} \geq q_{0,u} - \epsilon$. Let $\Theta = \bigcup_{m=1}^{M} \Theta_m \setminus \Theta_{g(\boldsymbol{\theta}_0)}$. Thus $\forall n \geq \max(N_1, N_2)$,

$$f_0(\boldsymbol{q}_n, \boldsymbol{\theta}_n) \geq (1 - |S|\epsilon) \inf_{\boldsymbol{\theta}' \in \Theta} \sum_{u \in S} \frac{q_{0,u} - \epsilon}{1 - |S|\epsilon} D_u(\boldsymbol{\theta}_n || \boldsymbol{\theta}'). \tag{74}$$

By continuity of f^*, we get

$$\liminf_{n \to \infty} f_0(\boldsymbol{q}_n, \boldsymbol{\theta}_n) \geq \liminf_{n \to \infty} (1 - |S|\epsilon) \inf_{\boldsymbol{\theta}' \in \Theta} \sum_{u \in S} \frac{q_{0,u} - \epsilon}{1 - |S|\epsilon} D_u(\boldsymbol{\theta}_n || \boldsymbol{\theta}') \tag{75}$$

$$= (1 - |S|\epsilon) \inf_{\boldsymbol{\theta}' \in \Theta} \sum_{u \in S} \frac{q_{0,u} - \epsilon}{1 - |S|\epsilon} D_u(\boldsymbol{\theta}_0 || \boldsymbol{\theta}'). \tag{76}$$

By continuity of f and letting $\epsilon \to 0$, we get

$$\liminf_{n \to \infty} f_0(\boldsymbol{q}_n, \boldsymbol{\theta}_n) \geq \inf_{\boldsymbol{\theta}' \in \Theta} \sum_{u \in S} q_{0,u} D_u(\boldsymbol{\theta}_0 || \boldsymbol{\theta}') = f_0(\boldsymbol{q}_0, \boldsymbol{\theta}_0). \tag{77}$$

Let $\{\boldsymbol{\theta}_k'\} \subset \Theta$ such that

$$\sum_{u=1}^{|\mathcal{U}|} q_{0,u} D_u(\boldsymbol{\theta}_0 || \boldsymbol{\theta}_k') \to \inf_{\boldsymbol{\theta}' \in \Theta} \sum_{u=1}^{|\mathcal{U}|} q_{0,u} D_u(\boldsymbol{\theta}_0 || \boldsymbol{\theta}') = f_0(\boldsymbol{q}_0, \boldsymbol{\theta}_0). \tag{78}$$

Thus,

$$\limsup_{n\to\infty} f_0(\boldsymbol{q}_n, \boldsymbol{\theta}_n) \leq \limsup_{n\to\infty} \sum_{u=1}^{|\mathcal{U}|} q_{n,u} D_u(\boldsymbol{\theta}_n || \boldsymbol{\theta}'_k) \tag{79}$$

$$\leq \sum_{u=1}^{|\mathcal{U}|} q_{0,u} D_u(\boldsymbol{\theta}_0 || \boldsymbol{\theta}'_k). \tag{80}$$

Taking limit as $k \to \infty$, we get

$$\limsup_{n\to\infty} f_0(\boldsymbol{q}_n, \boldsymbol{\theta}_n) \leq f_0(\boldsymbol{q}_0, \boldsymbol{\theta}_0). \tag{81}$$

Hence, we conclude that f_0 is continuous everywhere on its domain. Continuity of $\boldsymbol{q}^*$ follows from Berge's maximum theorem.

Proposition 2. Let $\beta_0(n, \alpha)$ satisfy the following equation.

$$e^{\beta_0(n,\alpha)} = \frac{4e^{|\mathcal{U}|+1}}{\alpha |\mathcal{U}|^{|\mathcal{U}|}} \beta_0(n,\alpha)^{2|\mathcal{U}|} n (1 + \log(n))^{|\mathcal{U}|+2}. \tag{82}$$

An upper bound on $\beta_0(n, \alpha)$ is $\beta(n, \alpha)$ as given in (33). Consequently,

$$e^{\beta(n,\alpha)} \geq \frac{4e^{|\mathcal{U}|+1}}{\alpha |\mathcal{U}|^{|\mathcal{U}|}} \beta(n,\alpha)^{2|\mathcal{U}|} n (1 + \log(n))^{|\mathcal{U}|+2}. \tag{83}$$

Proof. This result follows from Theorem 1 in [15] and expressing (82) in terms of the Lambert W-function. □

APPENDIX B

In this Appendix, we prove the theorems stated in section IV. We first establish some asymptotic convergence results.

Proposition 3. Let $\boldsymbol{\theta} \in \bigcup_{m=1}^{M} \Theta_m$ be the state of nature vector of parameters. Then the following holds for the policy that never stops and uses the proposed policy's recommendation and control law

$$\boldsymbol{\theta}^*(n) \xrightarrow{\text{a.s.}} \boldsymbol{\theta}, \tag{84}$$

$$\hat{r}(n) \xrightarrow{\text{a.s.}} g(\boldsymbol{\theta}), \tag{85}$$

$$\hat{\boldsymbol{\theta}}(n) \xrightarrow{\text{a.s.}} \boldsymbol{\theta}, \tag{86}$$

$$\boldsymbol{q}^*(\hat{\boldsymbol{\theta}}(n)) \xrightarrow{\text{a.s.}} \boldsymbol{q}^*(\boldsymbol{\theta}), \tag{87}$$

$$\frac{N_u(n)}{n} \xrightarrow{\text{a.s.}} q_u^*(\boldsymbol{\theta}), \forall u \in \mathcal{U}, \tag{88}$$

$$\tag{89}$$

Proof. We have from Lemma 2 that $N_u(n) \xrightarrow{\text{a.s.}} \infty$. By the Strong Law of Large Numbers and continuity of $\dot{b}_u$, we get that $\forall u \in \mathcal{U}$

$$\frac{S_u(n)}{N_u(n)} \to \dot{A}_u(\theta_u) \text{ and } \dot{b}_u\left(\frac{S_u(n)}{N_u(n)}\right) \xrightarrow{\text{a.s.}} \theta_u. \tag{90}$$

Thus, (84) holds. Note that $\Delta(\boldsymbol{x}, \Theta_m)$ is continuous at every $\boldsymbol{x} \in \Omega$ for any $m \in \mathcal{M}$. Consequently, (84) implies that

$$\Delta\big(\boldsymbol{\theta}^*(n), \Theta_{g(\boldsymbol{\theta})}\big) \xrightarrow{\text{a.s.}} \Delta(\boldsymbol{\theta}, \Theta_{g(\boldsymbol{\theta})}) = 0, \text{ and} \tag{91}$$

$$\Delta(\boldsymbol{\theta}^*(n), \Theta_i) \xrightarrow{\text{a.s.}} \Delta(\boldsymbol{\theta}, \Theta_i) > 0, \forall i \neq g(\boldsymbol{\theta}). \tag{92}$$

Thus, (85) holds. Consequently, (86) follows from (84), (85), (91) and (39). Note that (87) holds due to (86) and continuity of $\boldsymbol{q}^*$ (proposition 1). Note that it follows from lemma 2 that $\forall u \in \mathcal{U}$,

$$\left| \frac{N_u(n)}{n} - \frac{1}{n} \sum_{k=0}^{n-1} q_u^*(\hat{\boldsymbol{\theta}}(k)) \right| \leq \frac{|\mathcal{U}|(1+\sqrt{n})}{n} \xrightarrow{\text{a.s.}} 0. \tag{93}$$

Using Cesaro's lemma and (87), we have that $\forall u \in \mathcal{U}$,

$$\frac{1}{n} \sum_{k=0}^{n-1} q_u^*(\hat{\boldsymbol{\theta}}(k)) \xrightarrow{\text{a.s.}} q_u^*(\boldsymbol{\theta}). \tag{94}$$

Thus, (88) follows from (93) and (94). □

Lemma 3. Let $\boldsymbol{\theta} \in \bigcup_{m=1}^{M} \Theta_m$ be the state of nature vector of parameters. Then the following holds for the policy that never stops and uses the proposed policy's recommendation and control law

$$\frac{Z_{g(\boldsymbol{\theta})}(n)}{n} \xrightarrow{\text{a.s.}} D^*(\boldsymbol{\theta}). \tag{95}$$

Proof. Let $\mathcal{E}$ be the event given by

$$\mathcal{E} = \left\{ \forall u \in \mathcal{U}, \frac{S_u(n)}{N_u(n)} \to \dot{A}_u(\theta_u) \bigcap \frac{N_u(n)}{n} \to q_u^*(\boldsymbol{\theta}) \right\}. \tag{96}$$

By the Strong Law of Large Numbers and (88), we have that $\mathbb{P}_{\boldsymbol{\theta}}[\mathcal{E}] = 1$.

Claim 1. $\forall i \in \mathcal{M}$,

$$\frac{1}{n} \log \frac{\prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}, u)}{\sup_{\boldsymbol{\theta}' \in \Theta_i} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)} \xrightarrow{\text{a.s.}} \inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u=1}^{|\mathcal{U}|} q_u^*(\boldsymbol{\theta}) D_u(\boldsymbol{\theta} || \boldsymbol{\theta}'). \tag{97}$$

Proof. Let $i \in \mathcal{M}$. Since $\log$ is continuous and increasing, we have

$$\frac{1}{n} \log \frac{\prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}, u)}{\sup_{\boldsymbol{\theta}' \in \Theta_i} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)} = \inf_{\boldsymbol{\theta}' \in \Theta_i} \frac{1}{n} \sum_{u=1}^{|\mathcal{U}|} \log \frac{p(\underline{Y}^u(n); \boldsymbol{\theta}, u)}{p(\underline{Y}^u(n); \boldsymbol{\theta}', u)} \tag{98}$$

$$= \inf_{\boldsymbol{\theta}' \in \Theta_i} \frac{1}{n} \sum_{u=1}^{|\mathcal{U}|} \theta_u S_u(n) - N_u(n) A_u(\theta_u) - [\theta'_u S_u(n) - N_u(n) A_u(\theta'_u)] \tag{99}$$

$$= \inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u=1}^{|\mathcal{U}|} \frac{N_u(n)}{n} \left[D_u(\boldsymbol{\theta} || \boldsymbol{\theta}') + (\theta_u - \theta'_u) \left(\frac{S_u(n)}{N_u(n)} - \dot{A}_u(\theta_u) \right) \right] \tag{100}$$

$$= \inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u=1}^{|\mathcal{U}|} \frac{N_u(n)}{n} D_u(\boldsymbol{\theta} || \boldsymbol{\theta}') + \boldsymbol{W}_n^T (\boldsymbol{\theta} - \boldsymbol{\theta}'), \tag{101}$$

where $\boldsymbol{W}_n$ is the $|\mathcal{U}|$-dimensional vector such that $W_{n,u} = \frac{N_u(n)}{n} \left(\frac{S_u(n)}{N_u(n)} - \dot{A}_u(\theta_u) \right)$. Note that $\exists \{\boldsymbol{\theta}_k\} \subset \Theta_i$ such that

$$\sum_{u=1}^{|\mathcal{U}|} q_u^*(\boldsymbol{\theta}) D_u(\boldsymbol{\theta} || \boldsymbol{\theta}_k) \to \inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u=1}^{|\mathcal{U}|} q_u^*(\boldsymbol{\theta}) D_u(\boldsymbol{\theta} || \boldsymbol{\theta}'). \tag{102}$$

Thus on $\mathcal{E}$, we get

$$\limsup_{n \to \infty} \frac{1}{n} \log \frac{\prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}, u)}{\sup_{\boldsymbol{\theta}' \in \Theta_i} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)} \leq \limsup_{n \to \infty} \sum_{u=1}^{|\mathcal{U}|} \frac{N_u(n)}{n} D_u(\boldsymbol{\theta} || \boldsymbol{\theta}_k) + \boldsymbol{W}_n^T (\boldsymbol{\theta} - \boldsymbol{\theta}_k) \tag{103}$$

$$= \sum_{u=1}^{|\mathcal{U}|} q_u^*(\boldsymbol{\theta}) D_u(\boldsymbol{\theta} || \boldsymbol{\theta}_k). \tag{104}$$

Note that this holds for all $k \in \mathbb{N}$. Thus, taking limit $k \to \infty$, we get

$$\limsup_{n \to \infty} \frac{1}{n} \log \frac{\prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}, u)}{\sup_{\boldsymbol{\theta}' \in \Theta_i} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)} \leq \inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u=1}^{|\mathcal{U}|} q_u^*(\boldsymbol{\theta}) D_u(\boldsymbol{\theta} || \boldsymbol{\theta}'). \tag{105}$$

Let $\boldsymbol{q}^*(\boldsymbol{\theta}) \in \mathcal{P}_S$ for some $S \in 2^{\mathcal{U}} \setminus \{\phi\}$, where $\mathcal{P}_S$ is as defined in (48), We then have,

$$\inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u=1}^{|\mathcal{U}|} \frac{N_u(n)}{n} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') + \boldsymbol{W}_n^T(\boldsymbol{\theta} - \boldsymbol{\theta}') \geq \inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u \in S} \frac{N_u(n)}{n} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') + \boldsymbol{W}_n^T(\boldsymbol{\theta} - \boldsymbol{\theta}'). \tag{106}$$

Let $\epsilon \in \left(0, \min_{u \in S} q_u^*(\boldsymbol{\theta})/2\right)$. Thus on $\mathcal{E}$, $\exists N \in \mathbb{N}$ such that $\forall u \in S$, $\forall n > N$, $\frac{N_u(n)}{n} \geq q_u^*(\boldsymbol{\theta}) - \epsilon > 0$. So $\forall n \geq N$, R.H.S of (106) is bounded below

$$\inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u \in S} (q_u^*(\boldsymbol{\theta}) - \epsilon) D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') + \boldsymbol{W}_n^T(\boldsymbol{\theta} - \boldsymbol{\theta}'). \tag{107}$$

Note that the function $l : \mathbb{R}^{|\mathbb{U}|} \to \mathbb{R}$ given by

$$l(\boldsymbol{w}) = \inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u \in S} (q_u^*(\boldsymbol{\theta}) - \epsilon) D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') + \boldsymbol{w}^T(\boldsymbol{\theta} - \boldsymbol{\theta}') \tag{108}$$

is well-defined and concave on $\mathbb{R}^{|\mathbb{U}|}$ and thus continuous at $\mathbf{0}$. Using this we get that, on $\mathcal{E}$,

$$\begin{aligned}
\liminf_{n \to \infty} \frac{1}{n} \log \frac{\prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}, u)}{\sup_{\boldsymbol{\theta}' \in \Theta_i} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)} &\geq \liminf_{n \to \infty} l(\boldsymbol{W}_n) && (109)\\
&= l(\mathbf{0}) && (110)\\
&= \inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u \in S} (q_u^*(\boldsymbol{\theta}) - \epsilon) D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') && (111)\\
&= (1 - |S|\epsilon) \inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u \in S} \frac{q_u^*(\boldsymbol{\theta}) - \epsilon}{1 - |S|\epsilon} D_u(\boldsymbol{\theta}||\boldsymbol{\theta}'). && (112)
\end{aligned}$$

Note that this holds for any $\epsilon \in \left(0, \min_{u \in S} q_u^*(\boldsymbol{\theta})/2\right)$. Taking limit as $\epsilon \to 0$ and by continuity of f from proposition 1, we get

$$\liminf_{n \to \infty} \frac{1}{n} \log \frac{\prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}, u)}{\sup_{\boldsymbol{\theta}' \in \Theta_i} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)} \geq \inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u=1}^{|\mathcal{U}|} q_u^*(\boldsymbol{\theta}) D_u(\boldsymbol{\theta}||\boldsymbol{\theta}'). \tag{113}$$

From (105) and (113), we get the desired claim. □

Now using claim 1, we get that

$$\frac{Z_{g(\boldsymbol{\theta})}(n)}{n} = \min_{i \neq g(\boldsymbol{\theta})} \frac{Z_{g(\boldsymbol{\theta}),i}(n)}{n} \tag{114}$$

$$= \min_{i \neq g(\boldsymbol{\theta})} \frac{1}{n} \log \frac{\sup\limits_{\boldsymbol{\theta}'' \in \Theta_{g(\boldsymbol{\theta})}} \prod\limits_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}'', u)}{\sup\limits_{\boldsymbol{\theta}' \in \Theta_i} \prod\limits_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)} \tag{115}$$

$$= \min_{i \neq g(\boldsymbol{\theta})} \left[\frac{1}{n} \log \frac{\prod\limits_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}, u)}{\sup\limits_{\boldsymbol{\theta}' \in \Theta_i} \prod\limits_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)} - \frac{1}{n} \log \frac{\prod\limits_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}, u)}{\sup\limits_{\boldsymbol{\theta}'' \in \Theta_{g(\boldsymbol{\theta})}} \prod\limits_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}'', u)} \right] \tag{116}$$

$$\overset{\text{a.s.}}{\to} \min_{i \neq g(\boldsymbol{\theta})} \left[\inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u=1}^{|\mathcal{U}|} q_u^*(\boldsymbol{\theta}) D_u(\boldsymbol{\theta} || \boldsymbol{\theta}') - \inf_{\boldsymbol{\theta}'' \in \Theta_{g(\boldsymbol{\theta})}} \sum_{u=1}^{|\mathcal{U}|} q_u^*(\boldsymbol{\theta}) D_u(\boldsymbol{\theta} || \boldsymbol{\theta}'') \right] \tag{117}$$

$$= \min_{i \neq g(\boldsymbol{\theta})} \inf_{\boldsymbol{\theta}' \in \Theta_i} \sum_{u=1}^{|\mathcal{U}|} q_u^*(\boldsymbol{\theta}) D_u(\boldsymbol{\theta} || \boldsymbol{\theta}') \tag{118}$$

$$= \inf_{\boldsymbol{\theta}' \in \bigcup\limits_{m=1}^{M} \Theta_m \setminus \Theta_{g(\boldsymbol{\theta})}} \sum_{u=1}^{|\mathcal{U}|} q_u^*(\boldsymbol{\theta}) D_u(\boldsymbol{\theta} || \boldsymbol{\theta}') \tag{119}$$

$$= D^*(\boldsymbol{\theta}). \tag{120}$$

□

We proceed to show that the proposed policy is a $\bar{\alpha}$-correct policy. Observe that from Proposition 3, we get that $Z(n)$ is at-least linear in n almost surely for large n. On the other hand, the threshold $\beta(n, \alpha)$ is $O(\log n)$. Therefore, the proposed policy stops in finite time almost surely. To prove that the error probability is bounded by α, we use a concentration type inequality tailored for single parameter exponential families (Refer to Appendix B for details). We now rigourously prove these claims in the next theorem.

Theorem 1. *The proposed policy is a $\bar{\alpha}$-correct policy.*

Proof. We first prove that the proposed policy described has a finite stopping rule almost surely. Let $g(\boldsymbol{\theta}) = i$. Consider the event $\mathcal{E} = \left\{ \frac{Z_i(n)}{n} \to D^*(\boldsymbol{\theta}) \right\}$. From Lemma 3, we have that this event is of probability 1, that is $\mathbb{P}_{\boldsymbol{\theta}}[\mathcal{E}] = 1$. Let $\alpha \in (0, 1)$. Let $\epsilon > 0$. On $\mathcal{E}$, $\exists N \in \mathbb{N}$ such that $\forall n > N$,

$$Z(n) \geq Z_i(n) \geq \frac{n D^*(\boldsymbol{\theta})}{(1 + \epsilon)}. \tag{121}$$

Consequently,

$$\tau = \inf\{n \in \mathbb{N} : Z(n) \geq \beta(n, \alpha)\} \tag{122}$$

$$\leq N \vee \inf\left\{ n \in \mathbb{N} : \frac{n D^*(\boldsymbol{\theta})}{(1 + \epsilon)} \geq \beta(n, \alpha) \right\} \tag{123}$$

$$\leq N \vee \inf\left\{ n \in \mathbb{N} : \frac{n D^*(\boldsymbol{\theta})}{(1 + \epsilon)} \geq v(n) + w(\alpha) \right\} \tag{124}$$

where v and w are as given in proposition 2. Note that $\lim\limits_{t \to \infty} v'(t) = 0$. Hence,

$$\inf\left\{ n \in \mathbb{N} : \frac{n D^*(\boldsymbol{\theta})}{(1 + \epsilon)} \geq v(n) + w(\alpha) \right\} < \infty. \tag{125}$$

Consequently, $\tau < \infty$. Since $\mathbb{P}_{\boldsymbol{\theta}}[\mathcal{E}] = 1$, we get $\mathbb{P}_{\boldsymbol{\theta}}[\tau < \infty] = 1$. We first establish an upper bound on $Z_{j,i}$ for any $j \in \mathcal{M}$

$$Z_{j,i}(n) = \log \frac{\sup\limits_{\boldsymbol{\theta}' \in \Theta_j} \prod\limits_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)}{\sup\limits_{\boldsymbol{\theta}'' \in \Theta_i} \prod\limits_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}'', u)} \tag{126}$$

$$\leq \log \frac{\sup\limits_{\boldsymbol{\theta}' \in \Omega} \prod\limits_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)}{\prod\limits_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}, u)} \tag{127}$$

$$= \sum_{u=1}^{|\mathcal{U}|} N_u(n) D(\theta_u^*(n) || \theta_u). \tag{128}$$

We now proceed to prove that error probability is bounded by chosen α.

$$\mathbb{P}_{\boldsymbol{\theta}}[\hat{m} \neq i] \leq \mathbb{P}_{\boldsymbol{\theta}}\left[\exists n \in \mathbb{N}, \min_{j \neq i} Z_{j,i}(n) \geq \beta(n, \alpha)\right] \tag{129}$$

$$\leq \mathbb{P}_{\boldsymbol{\theta}}\left[\exists n \in \mathbb{N}, \exists j \in \mathcal{M} \setminus i : Z_{j,i}(n) \geq \beta(n, \alpha)\right] \tag{130}$$

$$\leq \sum_{n=1}^{\infty} \mathbb{P}_{\boldsymbol{\theta}}\left[\sum_{u=1}^{|\mathcal{U}|} N_u(n) D(\theta_u^*(n) || \theta_u) \geq \beta(n, \alpha)\right] \tag{131}$$

$$\leq \sum_{n=1}^{\infty} 2e^{-\beta} \left(\frac{\beta \lceil \beta \log n \rceil}{|\mathcal{U}|}\right)^{|\mathcal{U}|} e^{|\mathcal{U}|+1} \tag{132}$$

$$\leq \alpha \sum_{n=1}^{\infty} \frac{1}{2n(1+\log n)^2} \tag{133}$$

$$\leq \alpha. \tag{134}$$

The inequality (132) follows from Theorem 4 and (133) follows from Proposition 2. □

Theorem 2 (Almost-sure upper bound)**.** *Let* $\boldsymbol{\theta} \in \bigcup\limits_{m=1}^{M} \Theta_m$*. The proposed policy satisfies*

$$\mathbb{P}_{\boldsymbol{\theta}}\left[\limsup_{\alpha \to 0} \frac{\tau}{|\log \alpha|} \leq \frac{1}{D^*(\boldsymbol{\theta})}\right] = 1. \tag{135}$$

Proof. Let $g(\boldsymbol{\theta}) = i$. Consider the event $\mathcal{E} = \left\{\frac{Z_i(n)}{n} \to D^*(\boldsymbol{\theta})\right\}$. From proposition 3, we have that this event is of probability 1, that is $\mathbb{P}_{\boldsymbol{\theta}}[\mathcal{E}] = 1$. Let $\alpha \in (0, 1)$. Let $\epsilon > 0$. On $\mathcal{E}$, $\exists N \in \mathbb{N}$ such that $\forall n > N$,

$$Z(n) \geq Z_i(n) \geq \frac{nD^*(\boldsymbol{\theta})}{(1+\epsilon)}. \tag{136}$$

Consequently,

$$\tau = \inf\{n \in \mathbb{N} : Z(n) \geq \beta(n, \alpha)\} \tag{137}$$

$$\leq N \vee \inf\left\{n \in \mathbb{N} : \frac{nD^*(\boldsymbol{\theta})}{(1+\epsilon)} \geq \beta(n, \alpha)\right\} \tag{138}$$

$$\leq N \vee \inf\left\{n \in \mathbb{N} : \frac{nD^*(\boldsymbol{\theta})}{(1+\epsilon)} \geq v(n) + w(\alpha)\right\} \tag{139}$$

where v and w are as defined in (35) and (34) respectively. Note that $\forall t > 1, v'(t) > 0$ and $v''(t) < 0$. Also, $\lim\limits_{t \to \infty} v'(t) = 0$. Thus $\exists N_1 \in \mathbb{N}$ such that $\forall n \geq N_1$, $\frac{nD^*(\boldsymbol{\theta})}{(1+\epsilon)} > v(n)$. Also, $\exists N_2 \in \mathbb{N}$ such that $\forall n \geq N_2$, $v'(n) \in \overset{*}{-\frac{0.5D^*(\boldsymbol{\theta})}{(1+\epsilon)}, \frac{0.5D^*(\boldsymbol{\theta})}{(1+\epsilon)}}$. Let $N_3 = \max\{N, N_1, N_2\}$. Note that N_3 is not dependent on α. So, we get $\forall n \geq N_3$,

$$\tau \leq n + \frac{w(\alpha)}{\frac{D^*(\boldsymbol{\theta})}{(1+\epsilon)} - v'(n)}. \tag{140}$$

Consequently, $\forall n \geq N_3$,

$$\limsup_{\alpha \to 0} \frac{\tau}{|\log \alpha|} \leq \limsup_{\alpha \to 0} \frac{w(\alpha)}{\left[\frac{D^*(\boldsymbol{\theta})}{(1+\epsilon)} - v'(n)\right]|\log \alpha|}. \tag{141}$$

Note that $\lim_{\alpha \to 0} \frac{w(\alpha)}{|\log \alpha|} = 1$. This implies, $\forall n \geq N_3$,

$$\limsup_{\alpha \to 0} \frac{\tau}{|\log \alpha|} \leq \frac{1}{\frac{D^*(\boldsymbol{\theta})}{(1+\epsilon)} - v'(n)}. \tag{142}$$

Letting $n \to \infty$, we get

$$\limsup_{\alpha \to 0} \frac{\tau}{|\log \alpha|} \leq \frac{(1+\epsilon)}{D^*(\boldsymbol{\theta})}. \tag{143}$$

Now letting $\epsilon \to 0$, we get

$$\limsup_{\alpha \to 0} \frac{\tau}{|\log \alpha|} \leq \frac{1}{D^*(\boldsymbol{\theta})}. \tag{144}$$

□

Theorem 3 (Asymptotic optimality in expectation)**.** *Let* $\boldsymbol{\theta} \in \bigcup_{m=1}^{M} \Theta_m$*.. The proposed policy satisfies*

$$\limsup_{\alpha \to 0} \frac{\mathbb{E}_{\boldsymbol{\theta}}[\tau]}{|\log \alpha|} \leq \frac{1}{D^*(\boldsymbol{\theta})}. \tag{145}$$

Proof. Let $B_\xi(\boldsymbol{\theta}')$ denote the ξ-neighbourhood of $\boldsymbol{\theta}'$, that is, $B_\xi(\boldsymbol{\theta}') = \{\boldsymbol{\theta}'' : \|\boldsymbol{\theta}'' - \boldsymbol{\theta}'\| < \xi\}$ for $\xi > 0$. Let $\mathcal{I}_\xi(n)$ be the event given by

$$\mathcal{I}_\xi(n) \coloneqq \{\boldsymbol{\theta}^*(n) \in B_\xi(\boldsymbol{\theta})\}. \tag{146}$$

Let $g(\boldsymbol{\theta}) = m$ and $\boldsymbol{\theta} \in \Gamma_m^{(j)}$ for some $j \in [x_m]$. Since $\Gamma_m^{(j)}$ is open in $\text{aff}\Big(\Gamma_m^{(j)}\Big)$, $\exists \xi_0 > 0$ such that $B_{\xi_0}(\boldsymbol{\theta}) \cap \text{ aff}\Big(\Gamma_m^{(j)}\Big) \subset \Gamma_m^{(j)}$. Let $\xi_0' = \min_{m' \neq m \text{ or } i \neq j} \Delta(\boldsymbol{\theta}, \Gamma_{m'}^{(i)}) > 0$. Thus $\forall \xi \in \Big(0, \min(\xi_0, \frac{\xi_0'}{2})\Big)$,

$$\begin{aligned}
\mathcal{I}_\xi(n) &\implies \hat{r}(n) = m && (147)\\
&\implies \left\|\hat{\boldsymbol{\theta}}(n) - \boldsymbol{\theta}^*(n)\right\| \leq \rho \Delta(\boldsymbol{\theta}^*(n), \Theta_m) && (148)\\
&\implies \left\|\hat{\boldsymbol{\theta}}(n) - \boldsymbol{\theta}^*(n)\right\| \leq \rho \|\boldsymbol{\theta} - \boldsymbol{\theta}^*(n)\| && (149)\\
&\implies \left\|\hat{\boldsymbol{\theta}}(n) - \boldsymbol{\theta}^*(n)\right\| \leq \rho \xi && (150)\\
&\implies \left\|\hat{\boldsymbol{\theta}}(n) - \boldsymbol{\theta}^*(n)\right\| + \|\boldsymbol{\theta}^*(n) - \boldsymbol{\theta}\| \leq (1+\rho)\xi && (151)\\
&\implies \left\|\hat{\boldsymbol{\theta}}(n) - \boldsymbol{\theta}\right\| \leq (1+\rho)\xi. && (152)
\end{aligned}$$

From proposition 1 and (152), we get that given $\epsilon > 0$, $\exists \xi_1(\epsilon) \in \Big(0, \min(\xi_0, \frac{\xi_0'}{2})\Big)$ such that

$$\mathcal{I}_{\xi_1(\epsilon)}(n) \implies \max_{u \in \mathcal{U}} \Big| q_u^*(\hat{\boldsymbol{\theta}}(n)) - q_u^*(\boldsymbol{\theta}) \Big| < \epsilon. \tag{153}$$

Let $N \in \mathbb{N}$ and the event

$$\mathcal{E}_N(\epsilon) = \bigcap_{n=N^{1/4}}^{N} \mathcal{I}_{\xi_1(\epsilon)}(n). \tag{154}$$

The following claim is a consequence of the 'forced exploration' by the control law which ensures that each control is chosen at least around $\sqrt{n}$ times at time n.

Claim 2. $\exists K, C$ which are constants that depend on ϵ and $\boldsymbol{\theta}$ such that $\forall N \geq N' = 3^4 |\mathcal{U}|^8 + 1$,

$$\mathbb{P}_{\boldsymbol{\theta}}[\mathcal{E}_N^c(\epsilon)] \leq KN \exp\Big(-CN^{1/8}\Big). \tag{155}$$

Proof. Let $N \geq N'$. Thus, $\forall n \in [\sqrt{N}, N] \cap \mathbb{N}$, we get $\forall u \in \mathcal{U}, N_u(n) \geq \sqrt{n + |\mathcal{U}|^2} - 2|\mathcal{U}| > 0$. Note that

$$\mathbb{P}_{\boldsymbol{\theta}}[\mathcal{E}_N^c(\epsilon)] \leq \sum_{n=N^{1/4}}^{N} \mathbb{P}_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}^*(n) \notin B_{\xi_1(\epsilon)}(\boldsymbol{\theta})\right] \tag{156}$$

$$\leq \sum_{n=N^{1/4}}^{N} \sum_{u=1}^{|\mathcal{U}|} \mathbb{P}_{\boldsymbol{\theta}}[\theta_u^*(n) \notin (\theta_u - \xi, \theta_u + \xi)], \tag{157}$$

for some $\xi > 0$. Using a union bound and Chernoff inequality, we get that $\forall u \in \mathcal{U}$

$$\mathbb{P}_{\boldsymbol{\theta}}[\theta_u^*(n) \leq \theta_u - \xi] = \mathbb{P}_{\boldsymbol{\theta}}\left[\theta_u^*(n) \leq \theta_u - \xi, N_u(n) \geq s(n) = \sqrt{n + |\mathcal{U}|^2} - 2|\mathcal{U}|\right] \tag{158}$$

$$\leq \sum_{k=s(n)}^{n} \mathbb{P}_{\boldsymbol{\theta}}[\theta_u^*(n) \leq \theta_u - \xi, N_u(n) = k] \tag{159}$$

$$\leq \sum_{k=s(n)}^{n} \exp(-kD_u(\boldsymbol{\theta} - \xi||\boldsymbol{\theta})) \tag{160}$$

$$\leq \frac{e^{-s(n)D_u(\boldsymbol{\theta}-\xi||\boldsymbol{\theta})}}{1 - e^{-D_u(\boldsymbol{\theta}-\xi||\boldsymbol{\theta})}}, \tag{161}$$

where $\boldsymbol{\theta} + x$ is a vector given by $(\theta_1 + x, \theta_2 + x, \ldots, \theta_{|\mathcal{U}|} + x)$ for any scalar x. Similarly, we get

$$\mathbb{P}_{\boldsymbol{\theta}}[\theta_a^*(n) \leq \theta_u + \xi] \leq \frac{e^{-s(n)D_u(\boldsymbol{\theta}+\xi||\boldsymbol{\theta})}}{1 - e^{-D_u(\boldsymbol{\theta}+\xi||\boldsymbol{\theta})}}. \tag{162}$$

Let

$$C = \min_{u \in \mathcal{U}} (D_u(\boldsymbol{\theta} - \xi||\boldsymbol{\theta}) \vee D_u(\boldsymbol{\theta} + \xi||\boldsymbol{\theta})) \tag{163}$$

and

$$K = \sum_{u=1}^{|\mathcal{U}|} \left(\frac{e^{2|\mathcal{U}|D_u(\boldsymbol{\theta}-\xi||\boldsymbol{\theta})}}{1 - e^{-D_u(\boldsymbol{\theta}-\xi||\boldsymbol{\theta})}} + \frac{e^{2|\mathcal{U}|)D_u(\boldsymbol{\theta}+\xi||\boldsymbol{\theta})}}{1 - e^{-D_u(\boldsymbol{\theta}+\xi||\boldsymbol{\theta})}} \right). \tag{164}$$

Thus we get,

$$\mathbb{P}_{\boldsymbol{\theta}}[\mathcal{E}_N^c(\epsilon)] \leq \sum_{n=N^{1/4}}^{N} K \exp\left(-C\sqrt{n + |\mathcal{U}|^2}\right) \tag{165}$$

$$\leq KN \exp\left(-C\sqrt{N^{1/4} + |\mathcal{U}|^2}\right) \tag{166}$$

$$\leq KN \exp\left(-C\sqrt{N^{1/4}}\right) \tag{167}$$

$$= KN \exp\left(-CN^{1/8}\right). \tag{168}$$

□

The next claim discusses the convergence of empirical proportions on $\mathcal{E}_N(\epsilon)$.

Claim 3. $\exists N_\epsilon$ such that for $N \geq N_\epsilon$, it holds that on $\mathcal{E}_N(\epsilon)$,

$$\forall n \in [\sqrt{N}, N] \cap \mathbb{N}, \ \max_{u \in \mathcal{U}} \left| \frac{N_u(n)}{n} - q_u^*(\boldsymbol{\theta}) \right| \leq 2\epsilon. \tag{169}$$

Proof. Using lemma 2 and (153), we get that on $\mathcal{E}_N(\epsilon)$, $\forall n \in [\sqrt{N}, N] \cap \mathbb{N}$ and $\forall u \in \mathcal{U}$,

$$\left|\frac{N_u(n)}{n} - q_u^*(\boldsymbol{\theta})\right| \leq \left|\frac{N_u(n)}{n} - \frac{1}{n}\sum_{k=0}^{n-1} q_u^*(\hat{\boldsymbol{\theta}}(k)\right| + \left|\frac{1}{n}\sum_{k=0}^{n-1} q_u^*(\hat{\boldsymbol{\theta}}(k)) - q_u^*(\boldsymbol{\theta})\right| \tag{170}$$

$$\leq \frac{|\mathcal{U}|(\sqrt{n}+1)}{n} + \frac{N^{1/4}}{n} + \frac{1}{n}\sum_{k=N^{1/4}}^{n-1}\left|q_u^*(\hat{\boldsymbol{\theta}}(k)) - q_u^*(\boldsymbol{\theta})\right| \tag{171}$$

$$\leq \frac{2|\mathcal{U}|}{N^{1/4}} + \frac{1}{N^{1/4}} + \epsilon \tag{172}$$

$$= \frac{2|\mathcal{U}|+1}{N^{1/4}} + \epsilon \tag{173}$$

$$\leq 2\epsilon \tag{174}$$

when $N \geq \left(\frac{2|\mathcal{U}|+1}{\epsilon}\right)^4 = N_\epsilon$. □

Note that the GLRT statistic can be bounded below as follows.

$$Z(n) = \max_{i\in\mathcal{M}} \min_{j\neq i} Z_{i,j}(n) \tag{175}$$

$$\geq \min_{i\neq g(\boldsymbol{\theta})} Z_{g(\boldsymbol{\theta}),i}(n) \tag{176}$$

$$= \min_{i\neq g(\boldsymbol{\theta})} \log \frac{\sup_{\boldsymbol{\theta}'\in\Theta_{g(\boldsymbol{\theta})}} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}', u)}{\sup_{\boldsymbol{\theta}''\in\Theta_i} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}'', u)} \tag{177}$$

$$\geq \log \frac{\prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}, u)}{\sup_{\boldsymbol{\theta}''\in\Omega} \prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}'', u)} \tag{178}$$

$$= \log \frac{\prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}, u)}{\prod_{u=1}^{|\mathcal{U}|} p(\underline{Y}^u(n); \boldsymbol{\theta}^*(n), u)} \tag{179}$$

$$= n\sum_{u=1}^{|\mathcal{U}|} \frac{N_u(n)}{n}\left[D_u(\boldsymbol{\theta}||\boldsymbol{\theta}^*(n)) + (\theta_u - \theta_u^*(n))\left(\frac{S_u(n)}{N_u(n)} - \dot{A}_u(\theta_u)\right)\right] \tag{180}$$

$$= n\sum_{u=1}^{|\mathcal{U}|} \frac{N_u(n)}{n}\left[D_u(\boldsymbol{\theta}||\boldsymbol{\theta}^*(n)) + (\theta_u - \theta_u^*(n))\left(\dot{A}_u(\theta_u^*(n)) - \dot{A}_u(\theta_u)\right)\right] \tag{181}$$

$$= np\left(\boldsymbol{\theta}^*(n), \left(\frac{N_u(n)}{n}\right)_{u=1}^{|\mathcal{U}|}\right), \tag{182}$$

where $p : \Omega \times \mathcal{P} \to \mathbb{R}$ is the function given by,

$$p(\boldsymbol{\theta}', \boldsymbol{q}) \coloneqq \sum_{u=1}^{|\mathcal{U}|} q_u\left[D_u(\boldsymbol{\theta}||\boldsymbol{\theta}') + (\theta_u - \theta_u')\left(\dot{A}_u(\theta_u') - \dot{A}_u(\theta_u)\right)\right]. \tag{183}$$

Let

$$C_\epsilon^* = \inf_{\substack{\boldsymbol{\theta}': \left\|\boldsymbol{\theta}'-\boldsymbol{\theta}\right\| \leq \xi_1(\epsilon) \\ \boldsymbol{q}: \|\boldsymbol{q}-\boldsymbol{q}^*(\boldsymbol{\theta})\| \leq 2\epsilon}} p(\boldsymbol{\theta}', \boldsymbol{q}). \tag{184}$$

By the definition of $\mathcal{I}_{\xi_1(\epsilon)}(n)$ and claim 3, for $N \geq N_\epsilon$, on the event $\mathcal{E}_N(\epsilon)$, it holds that $\forall n \in [\sqrt{N}, N] \cap \mathbb{N}$,

$$Z(n) \geq nC_\epsilon^*. \tag{185}$$

Let $N \geq N_\epsilon$. On the event $\mathcal{E}_N(\epsilon)$,

$$\min(\tau, N) \leq \sqrt{N} + \sum_{n=\sqrt{N}}^{N} \mathbb{1}_{\tau > n} \tag{186}$$

$$\leq \sqrt{N} + \sum_{n=\sqrt{N}}^{N} \mathbb{1}_{Z(n) \leq \beta(n,\alpha)} \tag{187}$$

$$\leq \sqrt{N} + \sum_{n=\sqrt{N}}^{N} \mathbb{1}_{nC_\epsilon^* \leq \beta(n,\alpha)} \tag{188}$$

$$\leq \sqrt{N} + \frac{\beta(N,\alpha)}{C_\epsilon^*}. \tag{189}$$

We define

$$N_0(\alpha) \coloneqq \inf\left\{ N \in \mathbb{N} : \sqrt{N} + \frac{\beta(N,\alpha)}{C_\epsilon^*} \leq N \right\}. \tag{190}$$

So $\forall N \geq \max(N_0(\alpha), N_\epsilon)$, on $\mathcal{E}_N(\epsilon)$, we get

$$\min(\tau, N) \leq N \tag{191}$$

which implies

$$\tau \leq N. \tag{192}$$

Thus $\forall N \geq \max(N', N_0(\alpha), N_\epsilon)$,

$$\mathcal{E}_N(\epsilon) \subseteq (\tau \leq N) \tag{193}$$

and consequently,

$$\mathbb{P}_{\boldsymbol{\theta}}[\tau > N] \leq \mathbb{P}_{\boldsymbol{\theta}}[\mathcal{E}_N^c] \leq KN \exp\left(-CN^{1/8}\right). \tag{194}$$

So, we can upper bound the expectation of stopping time as

$$\mathbb{E}_{\boldsymbol{\theta}}[\tau] \leq N_0(\alpha) + N' + N_\epsilon + \sum_{N=1}^{\infty} KN \exp\left(-CN^{1/8}\right). \tag{195}$$

We now upper bound $N_0(\alpha)$ as follows.

$$N_0(\alpha) \leq \inf\left\{ N \in \mathbb{N} : \sqrt{N} + \frac{v(N) + w(\alpha)}{C_\epsilon^*} \leq N \right\}, \tag{196}$$

where $v(n)$ and $w(\alpha)$ are as defined in (35) and (34) respectively. Let $v_1(t) = \sqrt{t} + \frac{v(t)}{C_\epsilon^*}$. Note that $\forall t > 1, v_1'(t) > 0$ and $v_1''(t) < 0$. Also, $\lim_{t\to\infty} v_1'(t) = 0$ Thus, $\exists N_1 \in \mathbb{N}$ such that $\forall n \geq N_1$, $n > v_1(n)$. Also, $\exists N_2 \in \mathbb{N}$ such that $\forall n \geq N_2$, $v_1'(n) \in (-\frac{1}{2}, \frac{1}{2})$. Let $N_3 = \max\{N_1, N_2\}$. Note that N_3 is independent of α. Thus, $\forall n \geq N_3$,

$$N_0(\alpha) \leq n + \frac{\frac{w(\alpha)}{C_\epsilon^*}}{1 - v_1'(n)} \tag{197}$$

Consequently $\forall n \geq N_3$,

$$\limsup_{\alpha\to 0} \frac{N_0(\alpha)}{|\log \alpha|} \leq \frac{1}{C_\epsilon^*(1 - v_1'(n))}, \tag{198}$$

since $\lim_{\alpha\to 0} \frac{w(\alpha)}{|\log\alpha|} = 1$. Letting $n \to \infty$ we get,

$$\limsup_{\alpha\to 0} \frac{N_0(\alpha)}{|\log \alpha|} \leq \frac{1}{C_\epsilon^*}. \tag{199}$$

Using this in the inequality (195) as $\alpha \to 0$ we get,

$$\limsup_{\alpha\to 0} \frac{\mathbb{E}_{\boldsymbol{\theta}}[\tau]}{|\log \alpha|} \leq \frac{1}{C_\epsilon^*}. \tag{200}$$

By the continuity of p, we get

$$\lim_{\epsilon\to 0} C_\epsilon^* = D^*(\boldsymbol{\theta}). \tag{201}$$

So letting $\epsilon \to 0$ we get,

$$\limsup_{\alpha \to 0} \frac{\mathbb{E}_{\boldsymbol{\theta}}[\tau]}{|\log \alpha|} \leq \frac{1}{D^*(\boldsymbol{\theta})}. \tag{202}$$

□

APPENDIX C

In this section we extend the result in Theorem 2 in [14], stated for Bernoulli distributions, to single-parameter exponential family distributions.

Lemma 4. Let $a > 0$, $L \geq 2$. Let $\boldsymbol{Z} \in \mathbb{R}^L$ be a random variable such that $\forall \boldsymbol{\zeta} \in (\mathbb{R}^+)^L$

$$\mathbb{P}[\boldsymbol{Z} \geq \boldsymbol{\zeta}] \leq \exp\left(-a \sum_{l=1}^{L} \zeta_l\right). \tag{203}$$

Then $\forall \delta \geq L/a$,

$$\mathbb{P}\left[\sum_{l=1}^{L} Z_l \geq \delta\right] \leq \left(\frac{a\delta e}{L}\right)^L e^{-a\delta}. \tag{204}$$

Lemma 5. For any $u \in \mathcal{U}$, let $1 \leq t_u \leq n$. Let $\eta > 0$. Let E be the event given by

$$E = \bigcap_{u \in \mathcal{U}} \{t_u \leq N_u(n) \leq (1+\eta)t_u\}. \tag{205}$$

Then for $\beta \geq (1+\eta)(|\mathcal{U}| + \log 2)$, we have

$$\mathbb{P}_{\boldsymbol{\theta}}\left[\mathbb{1}_E \sum_{u \in \mathcal{U}} N_u(n) D_u(\boldsymbol{\theta}^*(n)||\boldsymbol{\theta}) \geq \beta\right] \leq 2\left(\frac{\beta e}{|\mathcal{U}|}\right)^{|\mathcal{U}|} e^{-\frac{\beta}{(1+\eta)}}. \tag{206}$$

Proof. We shall show that $\forall \boldsymbol{\zeta} \in (\mathbb{R}^+)^{|\mathcal{U}|}$,

$$\mathbb{P}_{\boldsymbol{\theta}}\left[\bigcap_{u \in \mathcal{U}} \{\mathbb{1}_E N_u(n) D_u(\boldsymbol{\theta}^*(n)||\boldsymbol{\theta}) \geq \zeta_u\}\right] \leq 2\exp\left[-\frac{\sum_{u \in \mathcal{U}} \zeta_u}{(1+\eta)}\right]. \tag{207}$$

Let $\boldsymbol{\zeta} \in (\mathbb{R}^+)^{|\mathcal{U}|}$. Let $\boldsymbol{c}^{(1)}$ and $\boldsymbol{c}^{(2)}$ be such that $\forall u \in \mathcal{U}$, $c_u^{(1)} < c_u^{(2)}$ and

$$t_u(1+\eta) D_u(\boldsymbol{c}^{(1)}||\boldsymbol{\theta}) = t_u(1+\eta) D_u(\boldsymbol{c}^{(2)}||\boldsymbol{\theta}) = \zeta_u. \tag{208}$$

Note that $\forall u \in \mathcal{U}$,

$$\mathbb{1}_E N_u(n) D_u(\boldsymbol{\theta}^*(n)||\boldsymbol{\theta}) \geq \zeta_u \implies \mathbb{1}_E t_u(1+\eta) D_u(\boldsymbol{\theta}^*(n)||\boldsymbol{\theta}) \geq \zeta_u \tag{209}$$

$$\implies E \bigcap \left\{\left\{\theta_u^*(n) \geq c_u^{(2)}\right\} \cup \left\{\theta_u^*(n) \leq c_u^{(1)}\right\}\right\}. \tag{210}$$

Now for a fixed $\boldsymbol{\lambda}$, let

$$M(n) = \exp\left\{\sum_{u \in \mathcal{U}} \lambda_u S_u(n) - N_u(n)\kappa_u(\lambda_u)\right\}. \tag{211}$$

So $\forall n' \leq n$,

$$M(n') = M(n'-1) \prod_{u \in \mathcal{U}} \exp\left\{\mathbb{1}_{\{U_{n'-1}=u\}}[\lambda_u T_u(Y_{n'}) - \kappa_u(\lambda_u)]\right\}. \tag{212}$$

Since $\mathbb{1}_{\{U_{n'}=u\}}$ is $\mathcal{F}_{n'-1}$-measurable and $Y_{n'}$ is conditionally independent of $\mathcal{F}_{n'-1}$, we get

$$\mathbb{E}_{\boldsymbol{\theta}}[M(n')|\mathcal{F}_{n'-1}] = M(n'-1). \tag{213}$$

Hence, $M(n)$ is a martingale and $\mathbb{E}_{\boldsymbol{\theta}}[M(n)] = 1$. $\forall u \in \mathcal{U}$, set $\lambda_u^{(1)} = c_u^{(1)} - \theta_u < 0$ and $\lambda_u^{(2)} = c_u^{(2)} - \theta_u > 0$, so that for $i \in \{1,2\}$, $\lambda_u^{(i)} \dot{A}_u(c_u^{(i)}) - \kappa_u(\lambda_u^{(i)}) = D_u(\boldsymbol{c}^{(i)}||\boldsymbol{\theta})$. Let

$$M_i(n) = \exp\left\{\sum_{u \in \mathcal{U}} \lambda_u^{(i)} S_u(n) - N_u(n)\kappa_u(\lambda_u^{(i)})\right\}, \tag{214}$$

for $i \in \{1,2\}$. Hence,

$$\mathbb{P}_{\boldsymbol{\theta}}\left[E \bigcap_{u\in\mathcal{U}}\left\{\left\{\theta_u^*(n) \le c_u^{(1)}\right\} \cup \left\{\theta_u^*(n) \ge c_u^{(2)}\right\}\right\}\right] \tag{215}$$

$$\le \mathbb{P}_{\boldsymbol{\theta}}\left[E \bigcap_{u\in\mathcal{U}}\left\{\dot{A}_u(\theta_u^*(n)) \le \dot{A}_u(c_u^{(1)})\right\}\right] + \mathbb{P}_{\boldsymbol{\theta}}\left[E \bigcap_{u\in\mathcal{U}}\left\{\dot{A}_u(\theta_u^*(n)) \ge \dot{A}_u(c_u^{(2)})\right\}\right] \tag{216}$$

$$\le \sum_{i=1}^{2}\mathbb{P}_{\boldsymbol{\theta}}\left[\mathbb{1}_E \sum_{u\in\mathcal{U}}\lambda_u^{(i)}S_u(n) \ge \sum_{u\in\mathcal{U}}N_u(n)\lambda_u^{(i)}\dot{A}_u(c_u^{(i)})\right] \tag{217}$$

$$\le \sum_{i=1}^{2}\mathbb{P}_{\boldsymbol{\theta}}\left[\mathbb{1}_E \sum_{u\in\mathcal{U}}\lambda_u^{(i)}S_u(n) - N_u(n)\kappa_u(\lambda_u^{(i)}) \ge \sum_{u\in\mathcal{U}}N_u(n)\lambda_u^{(i)}\dot{A}_u(c_u^{(i)}) - N_u(n)\kappa_u(\lambda_u^{(i)})\right] \tag{218}$$

$$\le \sum_{i=1}^{2}\mathbb{P}_{\boldsymbol{\theta}}\left[\mathbb{1}_E M_i(n) \ge \exp\left(\sum_{u\in\mathcal{U}}N_u(n)\left(\lambda_u^{(i)}\dot{A}_u(c_u^{(i)}) - \kappa_u(\lambda_u^{(i)})\right)\right)\right] \tag{219}$$

$$= \sum_{i=1}^{2}\mathbb{P}_{\boldsymbol{\theta}}\left[\mathbb{1}_E M_i(n) \ge \exp\left(\sum_{u\in\mathcal{U}}N_u(n)D_u(\boldsymbol{c}^{(i)}||\boldsymbol{\theta})\right)\right] \tag{220}$$

$$\le \sum_{i=1}^{2}\mathbb{P}_{\boldsymbol{\theta}}\left[\mathbb{1}_E M_i(n) \ge \exp\left(\sum_{u\in\mathcal{U}}t_u D_u(\boldsymbol{c}^{(i)}||\boldsymbol{\theta})\right)\right] \tag{221}$$

$$\le \sum_{i=1}^{2}\frac{\mathbb{E}_{\boldsymbol{\theta}}[\mathbb{1}_E M_i(n)]}{\exp\left(\sum\limits_{u\in\mathcal{U}}t_u D_u(\boldsymbol{c}^{(i)}||\boldsymbol{\theta})\right)} \tag{222}$$

$$\le \sum_{i=1}^{2}\frac{\mathbb{E}_{\boldsymbol{\theta}}[M_i(n)]}{\exp\left(\sum\limits_{u\in\mathcal{U}}t_u D_u(\boldsymbol{c}^{(i)}||\boldsymbol{\theta})\right)} \tag{223}$$

$$= \sum_{i=1}^{2}\exp\left(-\sum_{u\in\mathcal{U}}t_u D_u(\boldsymbol{c}^{(i)}||\boldsymbol{\theta})\right) \tag{224}$$

$$= 2\exp\left(-\sum_{u\in\mathcal{U}}\frac{\zeta_u}{(1+\eta)}\right). \tag{225}$$

Thus,

$$\mathbb{P}_{\boldsymbol{\theta}}\left[\bigcap_{u\in\mathcal{U}}\{\mathbb{1}_E N_u(n)D_u(\boldsymbol{\theta}^*(n)||\boldsymbol{\theta}) \ge \zeta_u\}\right] \le 2\exp\left(-\sum_{u\in\mathcal{U}}\frac{\zeta_u}{(1+\eta)}\right). \tag{226}$$

Let $Z_u = \mathbb{1}_E N_u(n)D_u(\boldsymbol{\theta}^*(n)||\boldsymbol{\theta})$ and $a = \frac{1}{1+\eta}$. Note that we have $\forall \boldsymbol{\zeta} \in (\mathbb{R}^+)^{|\mathcal{U}|}$,

$$\mathbb{P}_{\boldsymbol{\theta}}[\boldsymbol{Z} \ge \boldsymbol{\zeta}] \le 2\exp\left(-a\sum_{u\in\mathcal{U}}\zeta_u\right). \tag{227}$$

Let $\boldsymbol{Z}', \boldsymbol{\zeta}'$ be such that $\forall u \in \mathcal{U}, Z'_u = Z_u - \frac{\log 2}{a|\mathcal{U}|}$ and $\zeta'_u = \zeta_u - \frac{\log 2}{a|\mathcal{U}|}$. Thus,

$$\mathbb{P}_{\boldsymbol{\theta}}[\boldsymbol{Z}' \ge \boldsymbol{\zeta}'] \le 2\exp\left[-a\sum_{u\in\mathcal{U}}\left(\zeta'_u + \frac{\log 2}{a|\mathcal{U}|}\right)\right] \tag{228}$$

$$= \exp\left(-a\sum_{u\in\mathcal{U}}\zeta'_u\right). \tag{229}$$

This holds for all $\boldsymbol{\zeta}'$ such that $\forall u \in \mathcal{U}$, $\zeta'_u \ge -\frac{\log 2}{a|\mathcal{U}|}$. Hence, applying lemma 4 we get that $\forall \delta \ge \frac{|\mathcal{U}|}{a}$,

$$\mathbb{P}_{\boldsymbol{\theta}}\left[\sum_{u\in\mathcal{U}}Z'_u \ge \delta\right] \le \left(\frac{a\delta e}{|\mathcal{U}|}\right)^{|\mathcal{U}|}e^{-a\delta}. \tag{230}$$

Thus, $\forall \delta \geq \frac{|\mathcal{U}|+\log 2}{a}$,

$$\mathbb{P}_{\boldsymbol{\theta}}\left[\sum_{u\in\mathcal{U}} Z_u \geq \delta\right] \leq \left(\frac{a\left(\delta - \frac{\log 2}{a}\right)e}{|\mathcal{U}|}\right)^{|\mathcal{U}|} e^{-a\left(\delta-\frac{\log 2}{a}\right)} \tag{231}$$

$$\leq 2\left(\frac{a\delta e}{|\mathcal{U}|}\right)^{|\mathcal{U}|} e^{-a\delta}. \tag{232}$$

Hence, we get that for any $\beta \geq (1+\eta)(|\mathcal{U}| + \log 2)$,

$$\mathbb{P}_{\boldsymbol{\theta}}\left[\mathbb{1}_E \sum_{u\in\mathcal{U}} N_u(n) D_u(\boldsymbol{\theta}^*(n)||\boldsymbol{\theta}) \geq \beta\right] \leq 2\left(\frac{\beta e}{|\mathcal{U}|}\right)^{|\mathcal{U}|} e^{-\frac{\beta}{(1+\eta)}}. \tag{233}$$

□

Theorem 4.

$$\mathbb{P}_{\boldsymbol{\theta}}\left[\sum_{u\in\mathcal{U}} N_u(n) D_u(\boldsymbol{\theta}^*(n)||\boldsymbol{\theta}) \geq \beta\right] \leq 2e^{-\beta}\left(\frac{\beta\lceil\beta\log n\rceil}{|\mathcal{U}|}\right)^{|\mathcal{U}|} e^{|\mathcal{U}|+1}, \tag{234}$$

for $\beta \geq |\mathcal{U}| + 1 + \log 2$.

Proof. Let $\beta \geq |\mathcal{U}| + 1 + \log 2$ and $\eta = \frac{1}{\beta-1}$. Let $L = \left\lceil \frac{\log n}{\log(1+\eta)} \right\rceil$. Let $\mathbb{L} = \{1, 2, \ldots, L\}^{|\mathcal{U}|}$. Let G be the event

$$G = \left\{\sum_{u\in\mathcal{U}} N_u(n) D_u(\boldsymbol{\theta}^*(n)||\boldsymbol{\theta}) \geq \beta\right\}. \tag{235}$$

Let $H_{\boldsymbol{l}}$ be the event

$$H_l = \bigcap_{u\in\mathcal{U}} \left\{(1+\eta)^{l_u-1} \leq N_u(n) \leq (1+\eta)^{l_u}\right\}, \tag{236}$$

for any $\boldsymbol{l} \in \mathbb{L}$. We have

$$G = \bigcup_{\boldsymbol{l}\in\mathbb{L}} G \cap H_{\boldsymbol{l}}. \tag{237}$$

Thus,

$$\mathbb{P}_{\boldsymbol{\theta}}[G] \leq \sum_{\boldsymbol{l}\in\mathbb{L}} \mathbb{P}_{\boldsymbol{\theta}}[G \cap H_{\boldsymbol{l}}]. \tag{238}$$

Note that since $\beta \geq |\mathcal{U}| + 1 + \log 2$ and $\eta = \frac{1}{\beta-1}$, we get $\beta \geq (1+\eta)(|\mathcal{U}| + \log 2)$. So, applying lemma 5, we get that $\forall \boldsymbol{l} \in \mathbb{L}$,

$$\mathbb{P}_{\boldsymbol{\theta}}[G \cap H_{\boldsymbol{l}}] \leq 2\left(\frac{\beta e}{|\mathcal{U}|}\right)^{|\mathcal{U}|} e^{-\frac{\beta}{(1+\eta)}}. \tag{239}$$

Since $|\mathbb{L}| = L^{|\mathcal{U}|}$,

$$\mathbb{P}_{\boldsymbol{\theta}}[G] \leq 2\left(\frac{L\beta e}{|\mathcal{U}|}\right)^{|\mathcal{U}|} e^{-\frac{\beta}{(1+\eta)}}. \tag{240}$$

Note that $\log(1+\eta) \geq 1 - \frac{1}{1+\eta} = \frac{1}{\beta}$. Hence,

$$\mathbb{P}_{\boldsymbol{\theta}}[G] \leq 2e^{-\beta}\left(\frac{\beta\lceil\beta\log n\rceil}{|\mathcal{U}|}\right)^{|\mathcal{U}|} e^{|\mathcal{U}|+1}. \tag{241}$$

□

REFERENCES

[1] A. Deshmukh, S. Bhashyam, and V. V. Veeravalli, "Controlled sensing for composite multihypothesis testing with application to anomaly detection," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, Oct 2018, pp. 2109–2113.

[2] E. J. Horvitz, J. S. Breese, and M. Henrion, "Decision theory in expert systems and artificial intelligence," *International journal of approximate reasoning*, vol. 2, no. 3, pp. 247–302, 1988.

[3] H. Chernoff, "Sequential design of experiments," *Ann. Math. Statist.*, vol. 30, no. 3, pp. 755–770, 09 1959. [Online]. Available: http://dx.doi.org/10.1214/aoms/1177706205

[4] A. E. Albert, "The sequential design of experiments for infinitely many states of nature," *The Annals of Mathematical Statistics*, pp. 774–799, 1961.

[5] S. A. Bessler, "Theory and applications of the sequential design of experiments, k-actions and infinitely many experiments," Tech. Rep., 1960.

[6] S. Nitinawarat, G. K. Atia, and V. V. Veeravalli, "Controlled sensing for multihypothesis testing," *IEEE Transactions on Automatic Control*, vol. 58, no. 10, pp. 2451–2464, Oct 2013.

[7] M. Naghshvar and T. Javidi, "Active sequential hypothesis testing," *The Annals of Statistics*, vol. 41, no. 6, pp. 2703–2738, 2013.

[8] Y. Li, S. Nitinawarat, and V. V. Veeravalli, "Universal outlier hypothesis testing," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4066–4082, July 2014.

[9] K. Cohen and Q. Zhao, "Active hypothesis testing for anomaly detection," *IEEE Transactions on Information Theory*, vol. 61, no. 3, pp. 1432–1450, March 2015.

[10] N. K. Vaidhiyan and R. Sundaresan, "Learning to detect an oddball target," *IEEE Trans. Information Theory*, vol. 64, no. 2, pp. 831–852, 2018. [Online]. Available: https://doi.org/10.1109/TIT.2017.2778264

[11] G. R. Prabhu, S. Bhashyam, A. Gopalan, and R. Sundaresan, "Optimal odd arm identification with fixed confidence," *CoRR*, vol. abs/1712.03682, 2017. [Online]. Available: http://arxiv.org/abs/1712.03682

[12] E. Kaufmann, O. Cappé, and A. Garivier, "On the complexity of best-arm identification in multi-armed bandit models," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1–42, 2016.

[13] A. Garivier and E. Kaufmann, "Optimal best arm identification with fixed confidence," in *Conference on Learning Theory*, 2016, pp. 998–1027.

[14] S. Magureanu, R. Combes, and A. Proutiere, "Lipschitz bandits: Regret lower bound and optimal algorithms." in *COLT*, 2014, pp. 975–999.

[15] I. Chatzigeorgiou, "Bounds on the lambert function and their application to the outage analysis of user cooperation," *IEEE Communications Letters*, vol. 17, no. 8, pp. 1505–1508, August 2013.