LEARNING TO DETECT AN ANOMALOUS TARGET WITH OBSERVATIONS FROM AN EXPONENTIAL FAMILY

Gayathri R Prabhu, Srikrishna Bhashyam

Dept. of Electrical Engineering,
IIT Madras, Chennai 600036, India
Email: {ee15d035,skrishna}@ee.iitm.ac.in

Aditya Gopalan[1], Rajesh Sundaresan[12]

[1]Dept. of ECE and [2]Robert Bosch Centre
for Cyber-Physical Systems, IISc,
Bangalore 560012, India
Email: {aditya,rajeshs}@iisc.ac.in

ABSTRACT

The problem of identifying an anomalous arm from a set of K arms, with fixed confidence, is studied in a sequential decision-making scenario. Each arm's signal follows a distribution from the vector parameter exponential family. The actual parameters of the anomalous and regular arms are unknown. Further, the decision maker incurs a cost for switching from one arm to another. A sequential policy based on a modified generalised likelihood ratio statistic is proposed. The policy, with a suitable threshold, is shown to satisfy the given constraint on the probability of false detection. Further, the proposed policy is asymptotically optimal in terms of the total cost among all policies that satisfy the constraint on the probability of false detection.

Index Terms— conjugate prior, hypothesis testing, sequential analysis, search problems, switching costs.

1. INTRODUCTION

We consider the problem of detecting an anomalous arm from a set of K arms of a multi-armed bandit under a *fixed confidence* setting, i.e., with a constraint on the probability of false detection. Each arm follows a distribution from the vector exponential family parameterised by the natural vector parameter $\boldsymbol{\eta}$. As the name suggests, all arms except the "anomalous" one have the same parameter. The actual parameters of the anomalous and regular arms are unknown. At each successive stage or round, the decision maker chooses exactly one among the K arms for observation. The decision maker also incurs a cost whenever he switches from one arm to another. The goal is to minimise the overall cost of expected time for a reliable decision plus total switching cost, subject to a constraint on the probability of false detection. The above serves as a model of how one acquires data during a search task [1].

This work was supported by the Science and Engineering Research Board, Department of Science and Technology [grant no. EMR/2016/002503]. The authors acknowledge fruitful discussions with Aditya O. Deshmukh.

A commonly used test in such problems with unknown parameters is the generalised likelihood ratio test (GLRT) [2]. In our case, taking a cue from [3], we use a modified GLRT approach where the numerator of the statistic is replaced by an averaged likelihood function. The average is computed with respect to an artificial prior on the unknown parameters. The modified GLRT approach allows us to use a time invariant and a simple threshold policy that meets the constraint on probability of false detection.

Our interest in the exponential family is for three reasons.

- It unifies most of the widely used statistical models such as normal, Binomial, Poisson, and Gamma distributions.
- The generalisation forces us to rely on, and therefore bring out, the key properties of the exponential family that make the analysis tractable. These include the usefulness of the convex conjugate (or convex dual) of the log partition function, the existence of easily amenable formulae for relative entropy, and the usefulness of the conjugate prior in the analysis.
- The existence of conjugate priors enables extremely easy posterior updates. This is of great value in practice.

1.1. Related work

In [1], the authors have considered the anomalous arm identification problem with switching costs, but the statistics of the observations were assumed to be known and Poisson-distributed. In [3], the authors have considered a learning setting where the parameters of the Poisson distribution were not known but the switching costs were not taken into account. This work provides a significant generalisation of the results in [3] to the case of a general vector exponential family. This work also analyzes the effect of switching cost on search complexity in the presence of learning, thereby extending the results in [1] where the parameters were assumed known. For connections to, and limitations of, the works of Chernoff [4] and Albert [5], see [3, Sec. I-A]. A longer

version of our paper is available in [6].

1.2. Our contributions

- We provide a significant generalisation of the anomalous arm identification problem in [3], which dealt with the special case of Poisson observations, to the case of general vector exponential family observations.
- We modify the policy in [3] to incorporate switching costs based on the idea of slowed switching in [1], [7] and [8].
- We show that the proposed policy, which incorporates learning, is asymptotically optimal even with switching costs; the growth rate of the total cost, as the probability of false detection and the switching parameter are driven to zero, is the same as that without switching costs.
- We provide a method to verify an assumption that each arm is sampled at a nontrivial rate. Our rather general approach here, compared to [3], provides a simple proof of such a result for Poisson observations.

2. PRELIMINARIES

2.1. Vector exponential family basics

A probability distribution is a member of a vector exponential family if its probability density function (or probability mass function) can be written as

$$f\left(x|\boldsymbol{\eta}\right)=h\left(x\right)\exp\left(\boldsymbol{\eta}^{T}\mathbf{T}(x)-\mathcal{A}\left(\boldsymbol{\eta}\right)\right)\quad\forall x, \tag{1}$$

where $\boldsymbol{\eta}$ is the *natural* vector parameter of the family, $\boldsymbol{\eta}\in\mathbb{R}^{d}$ for some $d>0$ (or $\boldsymbol{\eta}$ is in some open convex subset of $\mathbb{R}^{d}$), $\mathbf{T}(x)\in\mathbb{R}^{d}$ is the sufficient statistic for the family, and $\mathcal{A}\left(\boldsymbol{\eta}\right):\boldsymbol{\eta}\rightarrow\mathbb{R}$ is a convex function known as the log partition function. The exponential family can also be parameterised using the *expectation* parameter defined as $\boldsymbol{\kappa}(\boldsymbol{\eta}):=E_{\boldsymbol{\eta}}[\mathbf{T}(x)]=\nabla_{\boldsymbol{\eta}}\mathcal{A}\left(\boldsymbol{\eta}\right)$ whenever $\mathcal{A}(\cdot)$ is continuously differentiable.

Define $\mathcal{F}(\boldsymbol{\kappa})$ as the convex conjugate of $\mathcal{A}(\boldsymbol{\eta})$ evaluated at an arbitrary $\boldsymbol{\kappa}$ given by

$$\mathcal{F}\left(\boldsymbol{\kappa}\right):=\sup_{\boldsymbol{\eta}\in\mathbb{R}^{d}}\{\boldsymbol{\eta}^{T}\boldsymbol{\kappa}-\mathcal{A}\left(\boldsymbol{\eta}\right)\}. \tag{2}$$

Since $\mathcal{A}(\cdot)$ is convex, we can recover $\mathcal{A}(\cdot)$ as the convex conjugate of $\mathcal{F}(\cdot)$. We assume henceforth that $\mathcal{F}(\cdot)$ and $\mathcal{A}(\cdot)$ are twice continuously differentiable at all points where they are finite. Optimising (2) over $\boldsymbol{\eta}$, we get that the optimising $\boldsymbol{\eta}$ satisfies $\boldsymbol{\kappa}(\boldsymbol{\eta})=\nabla_{\boldsymbol{\eta}}\mathcal{A}(\boldsymbol{\eta})$ which is the expectation parameter evaluated at $\boldsymbol{\eta}$. Similarly, optimizing $\mathcal{A}(\boldsymbol{\eta})$, we get $\boldsymbol{\eta}(\boldsymbol{\kappa})=\nabla_{\boldsymbol{\kappa}}\mathcal{F}(\boldsymbol{\kappa})$. Thus, the optimising $\boldsymbol{\eta}$ and $\boldsymbol{\kappa}$ are dual to each other and are in one-one correspondace. From [9, Section 3.3.2], we get

$$\mathcal{F}\left(\boldsymbol{\kappa}\right)+\mathcal{A}\left(\boldsymbol{\eta}\right)=\boldsymbol{\eta}^{T}\boldsymbol{\kappa} \tag{3}$$

The expressions for the Kullback-Leibler (KL) divergence or relative entropy in terms of the natural parameter and in terms of the expectation parameter using (3) are

$$\begin{aligned}D\left(\boldsymbol{\eta}_{1}||\boldsymbol{\eta}_{2}\right) &:= D\left(f(\cdot|\boldsymbol{\eta}_{1})||f(\cdot|\boldsymbol{\eta}_{2})\right)\\ &= \left(\boldsymbol{\eta}_{1}-\boldsymbol{\eta}_{2}\right)^{T}\boldsymbol{\kappa}_{1}-\mathcal{A}\left(\boldsymbol{\eta}_{1}\right)+\mathcal{A}\left(\boldsymbol{\eta}_{2}\right) \quad (4)\\ &= \left(\boldsymbol{\kappa}_{2}-\boldsymbol{\kappa}_{1}\right)^{T}\boldsymbol{\eta}_{2}+\mathcal{F}\left(\boldsymbol{\kappa}_{1}\right)-\mathcal{F}\left(\boldsymbol{\kappa}_{2}\right). \quad (5)\end{aligned}$$

2.2. Problem model

Let $K\geq 3$ be the number of arms available to the decision maker. Let the triplet $\psi=(i,\boldsymbol{\eta}_{1},\boldsymbol{\eta}_{2})$ denote the configuration of the arms, where the first component is the index of the anomalous arm, the second and the third components are the natural parameters of the anomalous and regular arms, respectively. We assume $\boldsymbol{\eta}_{1}\neq\boldsymbol{\eta}_{2}$. Let $\mathcal{P}\left(K\right)$ be the set of probability distributions on $\{1,2,\ldots,K\}$.

Let $\Pi\left(\alpha\right)$ be the set of admissible (desirable) policies that meet the following constraint on the probability of false detection:

$$\begin{aligned}\Pi\left(\alpha\right)=\{\pi:P\left(\delta\neq i|\psi=(i,\boldsymbol{\eta}_{1},\boldsymbol{\eta}_{2})\right)\leq\alpha,\\ \forall\psi\text{ such that }\boldsymbol{\eta}_{1}\neq\boldsymbol{\eta}_{2}\},\end{aligned} \tag{6}$$

with δ being the decision made when the algorithm stops. We define the stopping time of the policy as

$$\tau\left(\pi\right):=\inf\{n\geq 1:\overline{A}_{n}=(stop,\cdot)\}, \tag{7}$$

where $\overline{A}_{n}$ is the action taken by the policy at any stage n: $\overline{A}_{n}$ has two components $(continue,\cdot)$ or $(stop,\cdot)$. In the former case, the second argument indicates the arm to sample; in the latter case, it indicates the decision. The total cost will be the sum of the accumulated switching costs and the delay in arriving at a decision as in [7].

3. LOWER BOUND

A lower bound on the expected stopping time for any policy that satisfies the constraint on probability of false detection for the anomalous arm detection problem is given in the following proposition.

Proposition 1. *Fix α with $0<\alpha<1$. Let $\psi=(i,\boldsymbol{\eta}_{1},\boldsymbol{\eta}_{2})$ be the true configuration. For any $\pi\in\Pi\left(\alpha\right)$, we have*

$$E\left[\tau|\psi\right]\geq\frac{d_{b}\left(\alpha,1-\alpha\right)}{D^{*}\left(i,\boldsymbol{\eta}_{1},\boldsymbol{\eta}_{2}\right)} \tag{8}$$

where $d_{b}\left(\alpha,1-\alpha\right)$ is the binary relative entropy function and $D^{*}\left(i,\boldsymbol{\eta}_{1},\boldsymbol{\eta}_{2}\right)$ is defined as

$$\begin{aligned}D^{*}\left(i,\boldsymbol{\eta}_{1},\boldsymbol{\eta}_{2}\right) = \max_{0\leq\lambda(i)\leq 1}\Big[&\lambda\left(i\right)D\left(\boldsymbol{\eta}_{1}||\tilde{\boldsymbol{\eta}}\right)+\\ &\left(1-\lambda\left(i\right)\right)\frac{K-2}{K-1}D\left(\boldsymbol{\eta}_{2}||\tilde{\boldsymbol{\eta}}\right)\Big],\end{aligned} \tag{9}$$

with

$$\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta}(\tilde{\boldsymbol{\kappa}}) \text{ and } \tilde{\boldsymbol{\kappa}} = \frac{\lambda(i)\boldsymbol{\kappa}_1 + (1-\lambda(i))\frac{K-2}{K-1}\boldsymbol{\kappa}_2}{\lambda(i) + (1-\lambda(i))\frac{K-2}{K-1}}. \quad (10)$$

Also, the λ that maximises the expression in (9) is of the form

$$\lambda^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)(j) = \begin{cases} \lambda^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)(i), & \text{if } j = i \\ \frac{1-\lambda^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)(i)}{K-1}, & \text{if } j \neq i. \end{cases} \quad (11)$$

As the constraint on the probability of false detection approaches zero, $\alpha \to 0$, we have $d_b(\alpha, 1-\alpha)/\log(\alpha) \to -1$. Hence, we get that the conditional expected stopping time of the optimal policy scales at least as $-\log(\alpha)/D^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)$. The quantity $D^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)$ thus characterises the "complexity" of the learning problem at $(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)$. A proof of the result may be found in the longer version of our paper [6].

Corollary 2. *We have* $E[C(\pi)|\psi] \geq \frac{d_b(\alpha,1-\alpha)}{D^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)}$.

Proof. With the switching costs added, we have $C(\pi) \geq \tau(\pi)$, and the corollary follows from Proposition 1. □

A closed form expression for $\lambda^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)$ is not yet available. Hence, we make the following assumption.

Assumption 3. *Fix $K \geq 3$. Let λ^* maximise (9). There exists a constant $c_K \in (0,1)$, independent of $(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)$ but dependent on K, such that $\lambda^*(k,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)(j) \geq c_K > 0$ for all $j \in 1,2,\ldots,K$ and for all $(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)$ such that $\boldsymbol{\eta}_1 \neq \boldsymbol{\eta}_2$.*

In [6], we show that the assumption holds true for a wide range of members from the exponential family. The assumption suggests that a policy based on $\lambda^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)$ samples each arm at least c_K fraction of time independent of the ground truth. In case Assumption 3 does not hold, we could use a sampling policy with a forced exploration component as in [10] along with the modified GLRT approach.

4. PROPOSED POLICY

Let N_j^n denote the number of times the arm j was chosen for observation up to time n, i.e., $N_j^n = \sum_{t=1}^n 1_{\{A_t=j\}}$, where A_t is the arm chosen at time t. Clearly, we also have $n = \sum_{j=1}^K N_j^n$. Let $\mathbf{Y}_j^n$ denote the sum of the sufficient statistic of arm j up to time n, i.e.,$\mathbf{Y}_j^n = \sum_{t=1}^n \mathbf{T}(X_t)1_{\{A_t=j\}}$. Let $\mathbf{Y}^n$ denote the total sum of the sufficient statistic of all arms up to time n, i.e., $\mathbf{Y}^n = \sum_{j=1}^K \mathbf{Y}_j^n$.

When the parameters are unknown, a natural conjugate prior on $\boldsymbol{\eta}_1(j)$ and $\boldsymbol{\eta}_2(j)$ enables easy updates of the posterior distribution based on observations. The conjugate prior is taken to be a product distribution with each marginal once again coming from an exponential family of the same form and characterised by the *hyper-parameters* $\boldsymbol{\tau}$ and n_0, i.e.,

$$\begin{aligned} &f(\psi = (j,\boldsymbol{\eta}_1(j),\boldsymbol{\eta}_2(j))|H=j) \\ &= \mathcal{H}(\boldsymbol{\tau},n_0)\exp\{\boldsymbol{\tau}^T\boldsymbol{\eta}_1(j) - n_0\mathcal{A}(\boldsymbol{\eta}_1(j))\} \\ &\quad \times\mathcal{H}(\boldsymbol{\tau},n_0)\exp\{\boldsymbol{\tau}^T\boldsymbol{\eta}_2(j) - n_0\mathcal{A}(\boldsymbol{\eta}_2(j))\}, \quad (12) \end{aligned}$$

for a suitable normalisation $\mathcal{H}(\boldsymbol{\tau},n_0)$.

4.1. Modified GLR statistic

The modified GLR is defined as

$$\begin{aligned} Z_{ij}(n) &:= \log\frac{\tilde{f}(X^n,A^n|H=i)}{\hat{f}(X^n,A^n|H=j)} \quad (13) \\ &= \log\left\{\frac{\mathcal{H}(\boldsymbol{\tau},n_0)}{\mathcal{H}(\mathbf{Y}_i^n+\boldsymbol{\tau},N_i^n+n_0)}\right\} \\ &\quad + \log\left\{\frac{\mathcal{H}(\boldsymbol{\tau},n_0)}{\mathcal{H}(\mathbf{Y}^n-\mathbf{Y}_i^n+\boldsymbol{\tau},n-N_i^n+n_0)}\right\} \\ &\quad -\hat{\boldsymbol{\eta}}_1^T(j)\mathbf{Y}_j^n + N_j^n\mathcal{A}(\hat{\boldsymbol{\eta}}_1(j)) - \hat{\boldsymbol{\eta}}_2(j)^T(\mathbf{Y}^n-\mathbf{Y}_j^n) \\ &\quad + (n-N_j^n)\mathcal{A}(\hat{\boldsymbol{\eta}}_2(j)), \quad (14) \end{aligned}$$

where $\tilde{f}$ is the average likelihood function at time n, averaged over the conjugate prior in (12) over all configurations with $H=i$ and $\hat{f}$ is the maximum likelihood of observations till time n under $H=j$. Let $Z_i(n) := \min_{j\neq i} Z_{ij}(n)$ denote the modified GLR of i against its nearest alternative.

4.2. The policy $\pi_{SM}(L,\gamma)$

Fix $L \geq 1$ and $0 < \gamma \leq 1$. We now define the 'Sluggish, Modified GLR' policy $\pi_{SM}(L,\gamma)$ as follows:

At time n:

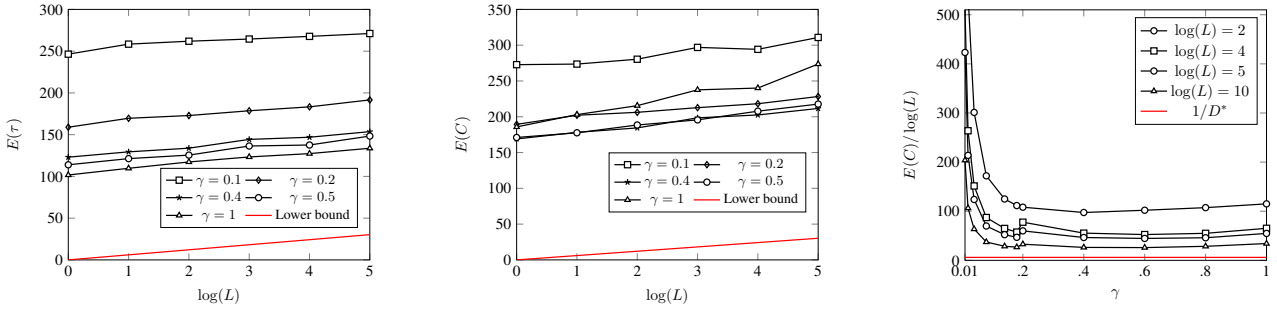
- Let $i^*(n) = \arg\max_i Z_i(n)$, an arm with the largest modified GLR at time n. Resolve ties uniformly at random.
- If $Z_{i^*(n)} < \log((K-1)L)$ then choose A_{n+1} via:
  - Generate U_{n+1}, a Bernoulli(γ) random variable independent of all other random variables with $\gamma > 0$.
  - If $U_{n+1} = 0$, then $A_{n+1} = A_n$.
  - If $U_{n+1} = 1$, then sample A_{n+1} according to the probability distribution $\lambda^*(i^*(n),\hat{\boldsymbol{\eta}}_1^n(i^*(n)),\hat{\boldsymbol{\eta}}_2^n(i^*(n)))$.
- If $Z_{i^*(n)} \geq \log((K-1)L)$ stop and declare $i^*(n)$ as the anomalous arm location.

5. ANALYSIS OF THE PROPOSED POLICY

The steps in this section verify that the proposed policy stops in finite time (Proposition 4), belongs to the desired set of policies (Proposition 5), and is asymptotically optimal (Proposition 6). Detailed proofs of these results are provided in the longer version of this paper [6, Appendix B]

Proposition 4. *(Stoppage) Fix the threshold parameter $L \geq 1$. Policy $\pi_{SM}(L,\gamma)$ stops in finite time with probability 1, that is, $P(\tau(\pi_{SM}(L,\gamma)) < \infty) = 1$.*

Proposition 5. *(Admissibility) Fix α. Let $L = 1/\alpha$. We then have $\pi_{SM}(L,\gamma) \in \Pi(\alpha)$.*



(a) (b) (c)

Fig. 1: Performance of $\pi_{SM}(L,\gamma)$ for Gaussian distribution with unknown means and unknown variances. $\mu_1 = 0$, $\sigma_1^2 = 2$, $\mu_2 = 1$, $\sigma_2^2 = 10$, $K = 8$, $g_{max} = 1$ and $D^* = 0.1653$.

Proposition 6. *(Achievability) Consider policy $\pi_{SM}(L,\gamma)$. Let $\psi = (i, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ be the true configuration. Then,*

$$\limsup_{L\to\infty} \frac{\tau(\pi_{SM}(L,\gamma))}{\log(L)} \leq \frac{1}{D^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)} \text{ a.s.,} \quad (15)$$

$$\limsup_{L\to\infty} \frac{E[\tau(\pi_{SM}(L,\gamma))|\psi]}{\log(L)} \leq \frac{1}{D^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)}, \quad (16)$$

and, furthermore,

$$\limsup_{L\to\infty} \frac{E[C(\pi_{SM}(L,\gamma))|\psi]}{\log(L)} \leq \frac{1+g_{max}\gamma}{D^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)}. \quad (17)$$

With these ingredients, we next state the main achievability result.

Theorem 7. *Consider K arms with configuration $\psi = (i, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$. Let $\left(\alpha^{(n)}\right)_{n\geq 1}$ be a sequence of tolerances such that $\lim_{n\to\infty} \alpha^{(n)} = 0$. Then, for each n, the policy $\pi_{SM}(L_n,\gamma)$ with $L_n = 1/\alpha^{(n)}$ belongs to $\Pi\left(\alpha^{(n)}\right)$. Furthermore,*

$$\liminf_{n\to\infty} \inf_{\pi\in\Pi\left(\alpha^{(n)}\right)} \frac{E[C(\pi)|\psi]}{\log(L_n)} \quad (18)$$

$$= \lim_{\gamma\downarrow 0} \lim_{n\to\infty} \frac{E[C(\pi_{SM}(L_n,\gamma))|\psi]}{\log(L_n)} = \frac{1}{D^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)}.$$

Here, we discuss the proofs briefly. We first show that under the true configuration, the test statistic has a positive drift and therefore crosses the threshold $\log((K-1)L)$ in finite time almost surely, thereby proving the result in Proposition 4. To prove admissibility in Proposition 5, we use elementary change of measure properties and the result that the policy stops and makes the decision when the statistic crosses the threshold. The use of the modified GLR statistic as opposed to the conventional GLR statistic simplifies this proof. For Proposition 6, we show that the positive drift of the statistic under the true configuration is equal to $D^*(i,\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)$. This relies on the convergence of the estimated parameters and the estimated anomalous arm location to the true values, something that is also proven along the way. The proof of Theorem 7 then follows from Propositions 1, 5 and 6.

6. SIMULATION RESULTS

Fig.1 shows (a) the empirical average delay versus $\log(L)$ for different values of γ, (b) the empirical average total cost (delay+switching costs) versus $\log(L)$ for different values of γ and (c) the ratio of empirical average total cost to $\log(L)$ versus γ averaged over 500 independent runs for the vector parameter Gaussian (both mean, variance unknown). The switching parameter in (a) and (b) is varied from 0.1, which corresponds to the sluggish implementation, to 1 when the policy switches according as per sampling strategy at each stage. As expected, we can observe that, in both (a) and (b), the slopes for policy match the slope of the lower bound, thereby validating the asymptotic optimality of the policy. In (c), as $\log(L) \to \infty$, the ratio of empirical average total cost to $\log(L)$ approaches the lower bound.

Also observe that in (a) for smaller values of γ, the average delay in arriving at a decision increases (due to low exploration) whereas, in (b) as γ decreases, the total cost decreases due to reduced switching (γ around 0.4 to 0.5 seems to be the best choice in this case). But, as $\gamma \downarrow 0$, policy becomes sluggish and requires more number of samples in making a decision, thereby resulting in an increased total cost as seen in (b). More simulation results for other distributions from the vector exponential family can be found in [6, Section VII].

7. CONCLUSION

We considered the problem of detecting an anomalous arm when the distributions are from a general vector exponential family. The parameters of the distributions are unknown. A sequential policy based on the modified GLR statistic was proposed. We showed that the proposed policy, which incorporates learning, is asymptotically optimal in terms of the total cost among all policies that satisfy a upper bound constraint on the probability of false detection.

8. REFERENCES

[1] Nidhin Vaidhiyan, Sripati P Arun, and Rajesh Sundaresan, "Neural dissimilarity indices that predict oddball detection in behaviour," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 4778–4796, 2017.

[2] H. Vincent Poor, *An Introduction to Signal Detection and Estimation (2Nd Ed.)*, Springer-Verlag New York, Inc., New York, NY, USA, 1994.

[3] Nidhin Koshy Vaidhiyan and Rajesh Sundaresan, "Learning to detect an oddball target," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 831–852, 2018.

[4] Herman Chernoff, "Sequential design of experiments," *The Annals of Mathematical Statistics*, vol. 30, no. 3, pp. 755–770, 1959.

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

[6] Gayathri R. Prabhu, Srikrishna Bhashyam, Aditya Gopalan, and Rajesh Sundaresan, "Learning to detect an oddball target with observations from an exponential family," *CoRR*, vol. abs/1712.03682, 2018.

[7] Nidhin Koshy Vaidhiyan and Rajesh Sundaresan, "Active search with a cost for switching actions," in *Information Theory and Applications Workshop (ITA), 2015*. IEEE, 2015, pp. 17–24.

[8] S Krishnaswamy, PT Akhil, A Arapostathis, S Shakkottai, and R Sundaresan, "Augmenting max-weight with explicit learning for wireless scheduling with switching costs," in *Proc. IEEE INFOCOM*, 2017, pp. 352–360.

[9] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

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