Sequential Experiment Design for Hypothesis Verification

Dhruva Kartik, Ashutosh Nayyar and Urbashi Mitra

Abstract—Hypothesis testing is an important problem with applications in target localization, clinical trials etc. Many active hypothesis testing strategies operate in two phases: an exploration phase and a verification phase. In the exploration phase, selection of experiments is such that a moderate level of confidence on the true hypothesis is achieved. Subsequent experiment design aims at improving the confidence level on this hypothesis to the desired level. In this paper, the focus is on the verification phase. A confidence measure is defined and active hypothesis testing is formulated as a confidence maximization problem in an infinite-horizon average-reward Partially Observable Markov Decision Process (POMDP) setting. The problem of maximizing confidence conditioned on a particular hypothesis is referred to as the hypothesis verification problem. The relationship between hypothesis testing and verification problems is established. The verification problem can be formulated as a Markov Decision Process (MDP). Optimal solutions for the verification MDP are characterized and a simple heuristic adaptive strategy for verification is proposed based on a zero-sum game interpretation of Kullback-Leibler divergences. It is demonstrated through numerical experiments that the heuristic performs better in some scenarios compared to existing methods in literature.

I. INTRODUCTION

Hypothesis testing is a classical problem and has been addressed in various settings. The problem can be described qualitatively as follows. An agent is interested in a phenomenon, and wants to test if the phenomenon conforms to any one of the hypotheses from a known class. The agent can perform various experiments and based on the observations from these experiments, it needs to infer the true hypothesis. As opposed to the one-shot hypothesis testing problem, an active agent can choose which experiment to perform based on the observations made in the past. The agent seeks to select experiments such that all false hypotheses are eliminated as quickly as possible.

Many active hypothesis testing strategies [1], [2] operate in two phases. The first phase is an exploration phase in which the experiment design is such that a moderate level of confidence is achieved on the true hypothesis. In most cases, this phase terminates in finite time almost surely [3]. The second is a verification phase in which the agent has a moderate level of confidence on some hypothesis and experiments are selected such that confidence on this hypothesis is improved to the desired level. When the desired confidence level is very high, the verification cost dominates the performance. In this paper, we make the notions of exploration and verification more formal and focus on analyzing the verification phase.

Active hypothesis testing finds applications in many areas such as sensor selection for target detection and localization, state tracking, design of clinical trials and learning unknown functions from queries [4]. Consequently, the verification phase plays an important role in all these applications.

We consider a slightly different mathematical formulation for hypothesis testing than previously explored [1], [2]. Using posterior belief on the set of hypotheses, we define a confidence level called Bayesian log-likelihood ratio. The objective is to design an experiment selection strategy that maximizes the expected rate of increase in the confidence level. Our contributions in this paper can be summarized as follows:

1) We formulate the verification problem as an infinite-horizon average-reward Markov Decision Process (MDP) problem.
2) We characterize the optimal rate using infinite-horizon Dynamic Programming (DP).
3) We identify a set of critical experiments. We then show that any strategy that selects these experiments while satisfying a stability criterion is asymptotically optimal.
4) We design a new heuristic experiment selection strategy and numerically show that it achieves better performance compared to existing methods in some scenarios.

The rest of the paper is organized as follows. In Section I-A, we discuss the relation between our problem and those in prior works. Section II formulates the problem. Section III relates the problem to the MDP framework and defines critical experiments. In Section IV, we solve the DP and in Section V, we describe an adaptive strategy and numerically compare it with existing policies. We conclude the paper in Section VI. Due to space constraints, some details are omitted. For detailed explanations, refer to [5].

A. Prior Work

The simplest active hypothesis testing problem was first formulated by Chernoff in [3] inspired by Wald’s analysis of the sequential probability ratio test [6]. Thereafter, it has been generalized in different ways depending on the target application [1], [2]. A major difference between our formulation and the formulation in these works is the reward structure. Prior works consider a combination of expected stopping time and Bayesian error probability. Fixed horizon problems have also been considered and they try to minimize the Bayesian error probability or maximal error probability [1]. We define a notion of confidence and maximize the expected rate of increase in confidence over long horizons. In prior formulations, if the
agent makes an error in guessing the true hypothesis, it incurs a cost of 1 (or some constant c) irrespective of its confidence level. Whereas in our formulation, we reward the agent for generating observations that result in a high confidence level on the true hypothesis. We believe that our formulation is related to the stopping time formulation because of the strong similarity in the results. In [3], [1], [2], the authors obtain asymptotically tight performance bounds and design policies that are asymptotically optimal. When the policies in these works are adapted to the verification problem defined herein, they turn out to be open-loop and randomized. A closed loop policy was designed in [7] but this may not always be asymptotically optimal. In this paper, we design a strategy for verification that is more adaptive and conjecture that it is asymptotically optimal.

B. Notation

Random variables/vectors are denoted by upper case boldface letters, their realization by the corresponding lower case letter. We use calligraphic fonts to denote sets (e.g. \(\mathcal{U}\)) and \(\Delta \mathcal{U}\) is the probability simplex over a finite set \(\mathcal{U}\). In general, subscripts are used as time index. There are two exceptions (\(\rho_j(n)\), \(X_j(n)\)) to this convention where the subscript denotes the hypothesis and \(n\) denotes time. For time indices \(n_1 \leq n_2\), \(Y_{n_1:n_2}\) is the short hand notation for the variables \((Y_{n_1}, Y_{n_1+1}, \ldots, Y_{n_2})\). For a strategy \(g\), we use \(\mathbb{P}^g(\cdot)\) and \(\mathbb{E}^g(\cdot)\) to indicate that the probability and expectation depend on the choice of \(g\). The Shannon entropy of a discrete distribution \(p\) over a finite space \(\mathcal{Y}\) is given by

\[
H(p) = -\sum_{y \in \mathcal{Y}} p(y) \log p(y).
\]  (1)

And the Kullback-Leibler divergence between distributions \(p\) and \(q\) is given by

\[
D(p||q) = \sum_{y \in \mathcal{Y}} p(y) \log \frac{p(y)}{q(y)}.
\]  (2)

II. Problem Formulation

Let \(\mathcal{G} \subset \mathbb{N}\) be a finite set of hypotheses and let \(\mathcal{H}\) be the true hypothesis. At each time \(n \in \mathbb{N}\), the agent can perform an experiment \(U_n \in \mathcal{U}\) and obtain an observation \(Y_n \in \mathcal{Y}\). For simplicity, let us also assume that the sets \(\mathcal{U}\) and \(\mathcal{Y}\) are finite. The observation \(Y_n\) at time \(n\) is given by

\[
Y_n = \xi(H, U_n, W_n),
\]  (3)

where \(\{ W_n : n \in \mathbb{N}\}\) is a collection of mutually independent and identically distributed primitive random variables. The probability of observing \(y\) after performing an experiment \(u\) under hypothesis \(h\) is denoted by \(p_h^u(y)\).

The information available at time \(n\), denoted by \(I_n\), is the collection of all experiments performed and the corresponding observations up to time \(n-1\), i.e.

\[
I_n = \{U_{1:n-1}, Y_{1:n-1}\}.
\]  (4)

Actions of the agent at time \(n\) can be functions of \(I_n\). Let the policy used for selecting the experiment be \(g_n\), i.e.

\[
U_n = g_n(I_n).
\]  (5)

The sequence of all the policies \(\{g_n\}\) is denoted by \(g\) which is referred to as a strategy. Let the collection of all such strategies be \(\mathcal{G}\).

Using the available information, the agent forms a posterior belief \(\rho(n)\) on \(\mathcal{H}\) at time \(n\) which is given by

\[
\rho_h(n) = \mathbb{P}[H = h | Y_{1:n-1}, U_{1:n-1}].
\]  (6)

Definition II.1 (Bayesian Log-Likelihood Ratio). The Bayesian log-likelihood ratio \(C_h(\rho)\) associated with an hypothesis \(h \in \mathcal{H}\) is defined as

\[
C_h(\rho) := \log \frac{\rho_h}{1 - \rho_h}.
\]  (7)

The Bayesian log-likelihood ratio (BLLR) is the logarithm of the ratio of the probability that hypothesis \(h\) is true versus the probability that hypothesis \(h\) is not true. BLLR is obtained by applying the logit function (also referred to as log-odds in statistics [8]) on the posterior belief \(\rho_h\). The logit function amplifies increments in \(\rho_h\) when \(\rho_h\) is close to 0 or 1. We can interpret BLLR as a measure of confidence on hypothesis \(h\) and thus, we refer to it as confidence level.

The objective is to design an experiment selection strategy \(g\) such that the confidence level \(C_H\) on the true hypothesis \(\mathcal{H}\) increases as quickly as possible. In other words, the total reward after acquiring \(N\) observations is the average rate of increase in the confidence level on the true hypothesis \(\mathcal{H}\) and is given by

\[
\frac{C_H(\rho(N+1)) - C_H(\rho(1))}{N}.
\]  (8)

More explicitly, we seek to design a strategy \(g\) that maximizes the asymptotic expected reward \(K(\rho)\) which is defined as

\[
K(\rho) := \lim_{N \to \infty} \inf \frac{1}{N} \mathbb{E}^g[ C_H(\rho(N+1)) - C_H(\rho(1))].
\]

Henceforth, we refer to this problem as the Expected Confidence Maximization (ECM) problem for hypothesis testing.

For a hypothesis \(h\) and a strategy \(g \in \mathcal{G}\), define \(J(g, h)\) as

\[
\lim_{N \to \infty} \inf \frac{1}{N} \mathbb{E}^g[ C_H(\rho(N+1)) - C_H(\rho(1)) | H = h].
\]

The value \(J(g, h)\) represents the performance of a strategy \(g\) conditioned on the hypothesis \(h\). Let

\[
J^*(h) = \sup_{g \in \mathcal{G}} J(g, h).
\]  (9)

For a given hypothesis \(h\), we refer to the problem of maximizing \(J(g, h)\) as the hypothesis verification problem. Let \(g^*(h)\) be an optimal verification strategy, i.e. it achieves the supremum in equation (9). We will later show that the existence of an optimal strategy \(g^*(h)\) is guaranteed under a mild assumption.
A. Hypothesis Testing vs Hypothesis Verification

The optimal verification cost $J^*(h)$ can be used to obtain an upper bound on the expected reward $K(g)$ in the hypothesis testing problem.

**Lemma II.1.** For any experiment selection strategy $g \in G$, we have

$$K(g) \leq \sum_{h \in H} \rho_h(1) J^*(h).$$

**Proof.** For any strategy $g \in G$, we have

$$K(g) = \sum_{h \in H} \rho_h(1) J(g, h) \leq \sum_{h \in H} \rho_h(1) J^*(h).$$

The last inequality follows from the definition of $J^*(h)$.

It is clear from the proof of Lemma II.1 that this upper bound is achieved by employing the strategy $g^*(h)$ when hypothesis $h$ is true. However, the agent cannot use different strategies under different hypotheses because it does not know the true hypothesis $H$. Therefore, we propose an experiment selection strategy of the following form. Similar strategies have also been used in [2].

$$\tilde{g}(\rho) = \begin{cases} g^*(h)(\rho) & \text{if for some } h, \rho_h > \tilde{\rho} \\ g^*(\rho) & \text{otherwise}, \end{cases}$$

where $0.5 < \tilde{\rho} < 1$ is a constant and $g^*$ is an exploration strategy. The interpretation of the strategy $\tilde{g}$ is that when the agent has a moderate level of confidence on some hypothesis $h$, it employs the corresponding verification strategy $g^*(h)$. This is to verify if hypothesis $h$ is indeed true by further improving its confidence level. When the agent is not very confident about any particular hypothesis, the agent employs an exploration strategy $g^*$. The primary purpose of the exploration strategy is to ensure that $\rho_H$ eventually crosses the threshold $\tilde{\rho}$. A naive exploration strategy is to randomly select every experiment uniformly. Better exploration strategies do exist [2], [7]. It remains to show that a strategy like $\tilde{g}$ can indeed achieve the upper bound in Lemma II.1. In this paper, we focus on the hypothesis verification problem. We derive sufficient conditions for an experiment selection strategy to be an optimal verification strategy.

III. **Markov Decision Process Formulation**

In this section, we show that the verification problem can be formulated as an infinite-horizon average-reward MDP problem. All of the following analysis is for $h = 1$ and with slight abuse of notation, we henceforth refer to $g^*(1)$ and $J(g, 1)$ as $g^*$ and $J(g)$, respectively. The same analysis can be repeated for any other $h$ to obtain similar results.

The state of the MDP is the posterior belief $\rho(n)$. The posterior belief is updated using Bayes’ rule. Thus, if $U_n = u$ and $Y_n = y$, we have

$$\rho_h(n + 1) = \frac{\rho_h(n) p_h^n(y)}{\sum_{h'} \rho_{h'}(n) p_{h'}^n(y)}.$$  

(13)

For convenience, we denote the Bayes’ update in (13) by

$$\rho(n + 1) = F(\rho(n), U_n, Y_n).$$

(14)

Since $H = 1$, we have $Y_n = \xi(1, U_n, W_n)$. Clearly, the dynamics of this system are Markovian. The expectation of average confidence rate under a strategy $g$ is given by

$$J_N(g) := \frac{1}{N} \mathbb{E}^g \left[ C_1(\rho(N + 1)) - C_1(\rho(1)) \right]$$

(15)

$$= \frac{1}{N} \mathbb{E}^g \left[ \sum_{n=1}^N [C_1(\rho(n + 1)) - C_1(\rho(n))] \right]$$

(16)

$$= \frac{1}{N} \mathbb{E}^g \left[ \sum_{n=1}^N r(\rho(n), U_n) \right].$$

(17)

Instantaneous reward for this MDP is the expected instantaneous increase in the confidence level and is given by

$$r(\rho, u) = \sum_{y \in Y} p^u_1(y) \log \frac{\rho_1 p^u_1(y)}{\sum_{j \neq 1} \rho_j p^u_j(y)} = \frac{1}{(1 - \rho_1)} \sum_{y \in Y} p^u_1(y) \log \frac{\rho_1 p^u_1(y)}{\sum_{j \neq 1} \rho_j p^u_j(y)},$$

(18)

where $\rho_j = \rho_j / (1 - \rho_1)$. Note that $\rho_j$ is a probability distribution over the set of alternate hypotheses $H = H \setminus \{1\}$. Also, notice that $r(\rho, u)$ is an KL-divergence between two distributions and hence, is always non-negative. The objective is to find a strategy $g^*$ that maximizes the following average reward

$$J(g) := \lim_{N \to \infty} \inf \frac{1}{N} \sum_{n=1}^N \mathbb{E}^g(r(\rho(n), U_n)).$$

(19)

We use Dynamic Programming (DP) to characterize optimal solutions for this infinite-horizon problem. In this framework, it can be shown that the randomized strategies used in [3], [1], [2] asymptotically achieve optimal rate $J^*$. Additionally, we identify a class of strategies that also achieve optimal rate and possibly, converge faster to the optimal rate than policies used in prior works.

Consider the following fixed point equation for the infinite horizon MDP

$$J' + w(\rho) = \max_u \left\{ r(\rho, u) + \sum_y p^u_1(y) w(F(\rho, u, y)) \right\},$$

(20)

where $J' \in \mathbb{R}$ is some constant and $w : \Delta H \to \mathbb{R}$ is some mapping. If such $J'$ and $w$ exist, then with some algebra (see [9] for details), we can conclude the following for any experiment selection strategy $g$ (possibly non-stationary)

$$\lim_{N \to \infty} \sup \frac{1}{N} \sum_{n=1}^N \mathbb{E}^g(r(\rho(n), U_n))$$

$$\leq \lim_{N \to \infty} \sup \frac{1}{N} (\mathbb{E}^g w(\rho(1)) - \mathbb{E}^g w(\rho(N + 1))) + J'$$

(22)

If we can show that

$$\lim_{N \to \infty} \sup \frac{1}{N} (\mathbb{E}^g w(\rho(1)) - \mathbb{E}^g w(\rho(N + 1))) \leq 0,$$

(23)

for every strategy $g$, then clearly the optimal rate $J^* \leq J'$. Additionally, if for some strategy $g^*$

$$\lim_{N \to \infty} \inf \frac{1}{N} (\mathbb{E}^{g^*} w(\rho(1)) - \mathbb{E}^{g^*} w(\rho(N + 1))) = 0$$

(24)
is satisfied and the experiment selected by $g^*$ is a maximizer in the fixed point equation (20), then $g^*$ is indeed an optimal strategy and $J^* = J'$ [9]. Our objective now is to find $J'$ and a function $w$ that satisfy these conditions. We make the following assumption on the conditional distributions $p_i^n(y)$. 

**Assumption 1.** There exists a constant $B > 0$ such that $|\lambda_j^i(u,y)| < B$ for every experiment $u$, observation $y$ and hypotheses $i, j \in H$, where

$$
\lambda_j^i(u,y) := \log \frac{p_i^n(y)}{p_j^n(y)}.
$$

We use the following quantities throughout our proofs. Let

$$
\alpha^* := \arg \max_{\alpha \in \Delta H} \min_{j \neq 1} \sum_u \alpha_u D(p_i^u||p_j^u) \quad (25)
$$

$$
\beta^* := \arg \min_{\beta \in \Delta H} \max_{j \neq 1} \sum_{j \neq 1} \beta_j D(p_i^u||p_j^u). \quad (26)
$$

Since the sets $U$ and $H$ are finite, existence of $\alpha^*$ and $\beta^*$ is guaranteed and also, by the minimax theorem

$$
\min_{\alpha \in \Delta U} \max_{j \neq 1} \sum_u \alpha_u D(p_i^u||p_j^u) = \max_{\beta \in \Delta H} \min_{j \neq 1} \sum_{j \neq 1} \beta_j D(p_i^u||p_j^u) =: R^*. \quad (27)
$$

We refer to the elements in the support of $\beta^*$ as critical hypotheses and those in the support of $\alpha^*$ as critical experiments. In particular, we show that the optimal rate $J^* = R^*$. 

**IV. Dynamic Programming Solution**

In this section, we solve the MDP formulated in Section III. Lemma IV.1 identifies a solution for the fixed point equation (20) and the subsequent Corollary IV.1 is used to obtain an upper bound on $J^*$. We then show that this upper bound can indeed be achieved.

**Lemma IV.1.** The fixed point equation (20) is satisfied with $J' = R^*$ and

$$
w(\rho) = - \sum_{j \neq 1} \beta_j^* \log \frac{\rho_j}{1 - \rho_j} = - \sum_{j \neq 1} \beta_j^* \log \tilde{\rho}_j. \quad (28)
$$

*Any critical experiment is a maximizer in the equation (20).*

**Proof.** We can simply verify the fixed point equation (20) by substitution. For details, see [5].

**Corollary IV.1.** For any strategy $g$, we have

$$
\lim_{N \to \infty} \sup \frac{1}{N} \left( \mathbb{E}^g w(\rho(1)) - \mathbb{E}^g w(\rho(N+1)) \right) \quad (29)
$$

$$
= \lim_{N \to \infty} \sup \frac{1}{N} \sum_{j \neq 1} \beta_j^* \mathbb{E}^g \log \tilde{\rho}_j(N + 1) \leq 0. \quad (30)
$$

**Proof.** This is simply because $\tilde{\rho}_j(N + 1) \leq 1$.

**Theorem IV.1.** The optimal average rate $J^* \leq R^*$.

**Proof.** This directly follows from the fact that $w$ defined in Lemma IV.1 satisfies inequality (23) and with $J' = R^*$, the fixed point equation (20) is satisfied.

**Theorem IV.2.** The optimal average rate $J^* = R^*$.

**Proof.** The main idea is to show using Strong Law of Large Numbers (SLLN) [10] that when experiments are selected independently using the distribution $\alpha^*$, the criterion (24) is satisfied with $w$ as defined in Lemma IV.1. Assumption 1 allows us to use SLLN. See [5] for the complete proof. 

To summarize, the following conditions are sufficient for a verification strategy $g$ to be asymptotically optimal:

1) The strategy $g$ only selects critical experiments, i.e. experiments from the support of $\alpha^*$.

2) The following stability criterion is satisfied, i.e.

$$
\lim_{N \to \infty} \frac{1}{N} \sum_{j \neq 1} \beta_j^* \mathbb{E}^g \log \tilde{\rho}_j(N + 1) = 0. \quad (31)
$$

These conditions suggest that there could be many strategies other than the open-loop randomized strategy used in Theorem IV.2 that achieve asymptotic optimality.

**V. Numerical Results**

In this section, we propose a new heuristic based on a Kullback-Leibler divergence zero-sum game and demonstrate numerically that this heuristic’s performance is close to the maximum achievable confidence rate $R^*$. We first briefly describe all the strategies used in our experiments.

**1) Extrinsic Jensen-Shannon (EJS) Divergence:** Extrinsic Jensen-Shannon divergence as a notion of information was first introduced in [7]. Using our notation, EJS for a query $u$ at some belief state $\rho$ is given by

$$
EJS(\rho, u) = \mathbb{E}[C(F(\rho, u, Y)) - C(\rho)], \quad (32)
$$

where the expectation of $Y$ is with respect to the distribution $\sum_{h \in H} \rho_h p_h^n(y)$ and

$$
C(\rho) = \sum_{i \in H} \rho_i \log \frac{\rho_i}{1 - \rho_i} = \sum_{i \in H} \rho_i C_i(\rho). \quad (33)
$$

The EJS heuristic selects the experiment $u$ that maximizes $EJS(\rho, u)$ for a given state $\rho$.

**2) Open Loop Verification (OPE):** As discussed earlier, the strategies in [2], [1], [3] when specialized to verification are open-loop and randomized. According to this strategy, the queries are randomly selected independently in an open-loop manner from the distribution $\alpha^*$. Recall that this strategy is asymptotically optimal as shown in Theorem IV.2.

**3) KL-divergence Zero-sum Game (KLZ):** We design the following heuristic. Consider a zero-sum game in which the first player (maximizing) selects an experiment $u \in U$ and the second player (minimizing) selects an alternate hypothesis $j \in H$. The payoff for this zero-sum game is the KL-divergence $D(p_i^n||p_j^n)$. The agent picks an experiment $u$ that maximizes

$$
\mathcal{S}(\rho, u) := \sum_{j \neq 1} \tilde{\rho}_j D(p_i^n||p_j^n). \quad (34)
$$

This strategy can be interpreted as the first player’s best-response when the second player uses the mixed strategy $\tilde{\rho}_j$ to select an alternate hypothesis. Note that the mixed strategy $\alpha^*$ used in OPE is a Nash equilibrium strategy for the maximizing player.
A. Simulation Setup

To simulate these heuristics, we first consider a simple setup with three hypotheses and two queries. The conditional distributions \( p_i^j(y) \) for each of these queries are illustrated in Figure 1. The queries are designed such that when \( H = \hat{h}_0 \), the agent is forced to make both queries \( u^1 \) and \( u^2 \). This is because hypotheses \( \hat{h}_0 \) and \( \hat{h}_2 \) are indistinguishable under query \( u^1 \) and similarly, hypotheses \( \hat{h}_0 \) and \( \hat{h}_1 \) are indistinguishable under query \( u^2 \). We illustrate the evolution of expected confidence rate \( J_N \) under hypothesis \( \hat{h}_0 \) in Figure 2. The heuristics EJS and KLZ come very close to the maximum achievable rate. OPE eventually achieves maximal rate but very slowly.

|        | \( y = 0 \) | \( y = 1 \) |        | \( y = 0 \) | \( y = 1 \) |
|--------|------------|------------|--------|------------|------------|
| \( h_0 \) | 0.8        | 0.2        | \( h_0 \) | 0.8        | 0.2        |
| \( h_1 \) | 0.2        | 0.8        | \( h_2 \) | 0.8        | 0.2        |
| \( h_2 \) | 0.8        | 0.2        |        |            |            |

(a) Query \( u^1 \) (b) Query \( u^2 \)

Fig. 1: Conditional distributions \( p_i^j(y) \) for each query

We also consider another setup in which EJS is not optimal and yet, our heuristic KLZ comes close to optimal performance. We refer the reader to [5] for more details.

B. Stopping Time Formulation

In [3], [1], [11], a stopping time formulation for hypothesis testing is considered. The sampling process stops when the belief on some hypothesis exceeds a threshold or equivalently, when the confidence \( C_h(\rho) > \log L \), where \( L \) is a parameter. Let this stopping time be \( N \). Under this stopping criterion, we numerically study the expected stopping time for all the strategies discussed. The plots in Figure 3 depict the quantity \( E[N] / \log L \) as a function of the parameter \( L \). Numerical results suggest that our heuristic performs better even in the stopping time formulation.

![Fig. 1](image1.jpg)

![Fig. 2](image2.jpg)

![Fig. 3](image3.jpg)

Fig. 2: Evolution of expected confidence rate \( J_N \) under hypothesis \( \hat{h}_0 \). Note the subpar performance of OPE.

VI. Conclusion

In this paper, we formulate the problem of quickly verifying a given hypothesis using observations from experiments as an infinite horizon average cost MDP. We characterize the optimal rate of this MDP using infinite horizon dynamic programming. A stability criterion arises out of the DP equations. We show that any strategy that satisfies this stability criterion while selecting experiments from a critical set is asymptotically optimal. We proposed a heuristic adaptive strategy and numerically demonstrated that it performs better than open-loop policies in the non-asymptotic regime. For future work, we intend to use this stability criterion to design strategies with better non-asymptotic performance.

REFERENCES

[1] Sirin Nitinawarat, George K Atia, and Venugopal V Veeravalli, “Controlled sensing for multi-hypothesis testing,” IEEE Transactions on Automatic Control, vol. 58, no. 10, pp. 2451–2464, 2013.
[2] Mohammad Naghshvar, Tara Javidi, et al., “Active sequential hypothesis testing,” The Annals of Statistics, vol. 41, no. 6, pp. 2703–2738, 2013.
[3] Herman Chernoff, “Sequential design of experiments,” The Annals of Mathematical Statistics, vol. 30, no. 3, pp. 755–770, 1959.
[4] Mohammad Naghshvar, Tara Javidi, and Kamalika Chaudhuri, “Bayesian active learning with non-persistent noise,” IEEE Transactions on Information Theory, vol. 61, no. 7, pp. 4080–4098, 2015.
[5] Dhruba Kartik, Ashutosh Nayyar, and Urbashi Mitra, “Asymptotically optimal sequential experimental design for hypothesis verification,” https://bit.ly/2DVsaa7.
[6] Abraham Wald, Sequential analysis, Courier Corporation, 1973.
[7] Mohammad Naghshvar and Tara Javidi, “Extrinsic jensen-shannon divergence with application in active hypothesis testing,” in Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on. IEEE, 2012, pp. 2191–2195.
[8] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant, Applied logistic regression, vol. 398, John Wiley & Sons, 2013.
[9] Punganamala Ramana Kumar and Pravin Varaiya, Stochastic systems: Estimation, identification, and adaptive control, vol. 75, SIAM, 2015.
[10] Rick Durrett, Probability: theory and examples, Cambridge university press, 2010.
[11] Mohammad Naghshvar and Tara Javidi, “Sequentiality and adaptivity gains in active hypothesis testing,” IEEE Journal of Selected Topics in Signal Processing, vol. 7, no. 5, pp. 768–782, 2013.