

Adaptive Sampling for Estimating Distributions: A Bayesian Upper Confidence Bound Approach

Dhruva Kartik
Neeraj Sood
Urbashi Mitra

University of Southern California, Los Angeles

Tara Javidi

University of California, San Diego

MOKHASUN@USC.EDU

NSOOD@USC.EDU

UBLI@USC.EDU

TJAVIDI@UCSD.EDU

Abstract

The problem of adaptive sampling for estimating probability mass functions (pmf) uniformly well is considered. Performance of the sampling strategy is measured in terms of the worst-case mean squared error. A Bayesian variant of the existing upper confidence bound (UCB) based approaches is proposed. It is shown analytically that the performance of this Bayesian variant is no worse than the existing approaches. The posterior distribution on the pmfs in the Bayesian setting allows for a tighter computation of upper confidence bounds which leads to significant performance gains in practice. Using this approach, adaptive sampling protocols are proposed for estimating SARS-CoV-2 seroprevalence in various groups such as location and ethnicity. The effectiveness of this strategy is discussed using data obtained from a seroprevalence survey in Los Angeles county.

Keywords: Adaptive sampling, Distribution estimation, Bayesian methods, Upper Confidence Bound

1. Introduction

We frequently encounter scenarios where we need to estimate a finite collection of pmfs from their samples. We have a limited sample budget and we would like to adaptively choose which pmf to sample from so that we have uniformly good estimates of all the pmfs. The *goodness* of our estimate can be evaluated using various distribution distance metrics. In this paper, we will focus on the mean squared error. A concrete application that fits this setting is that of estimating SARS-CoV-2 incidence rates in various geographic regions. Incidence rate, the fraction of *currently* infected individuals, is a crucial metric for assessing the safety of a given region. Incidence rate cannot be estimated in an unbiased way simply based on the reported cases, it needs to be estimated by conducting randomized tests in each region of interest. Typically, such randomized surveys can be conducted by a central authority which has a limited testing budget. A natural question that arises in this scenario is that of allocating the available tests to various regions in an efficient manner. We may also be interested in other parameters such as SARS-CoV-2 seroprevalence which is an indicator of an individual's potential immunity against the virus, and other kinds of groups like ethnicity, age etc. in addition to the geographic location. Other applications of this setting include dynamics estimation in Markov Decision Processes (MDPs) and text compression which are discussed in greater detail in (Shekhar et al., 2020).

When we are interested in minimizing the mean squared error for every pmf, an efficient sampling strategy would be to let the number of tests be proportional to the *variance* of the associated pmf. However, the key problem is that we do not know the variance of the pmf. Therefore, we need to adapt our allocation as we learn the distribution over time. An upper confidence bound (UCB) based approach was proposed in (Shekhar et al., 2020). The main idea in this approach is to first form an upper bound on the variance of each pmf. As a function of this upper bound and the number of samples acquired so far for each pmf, the next sample is chosen accordingly. It is shown that the number of samples obtained in this manner would approximately be the same as that of an *oracle* that allocates samples using the knowledge of the variance of the underlying distributions. While this strategy satisfies many such interesting theoretical properties, we observed that it tends to be more conservative than necessary in practice.

Our main goal in this paper is to devise a less conservative approach while ensuring the same performance guarantees as in (Shekhar et al., 2020). Our contributions towards this goal are (i) we assume a Dirichlet prior on the pmfs and compute aforementioned upper confidence bounds using the posterior belief on the pmfs. This, combined with a stronger set of inequalities gives us noticeably tighter bounds on the variance of the pmfs; (ii) we analytically prove that this Bayesian approach for computing confidence bounds is no worse than the one in (Shekhar et al., 2020); and (iii) we employ these methods for estimating SARS-CoV-2 seroprevalence in various categories using data obtained by a randomized survey (Sood et al., 2020) conducted in Los Angeles county. Our analysis of the survey data illustrates the importance of adaptive sampling in order to obtain better estimates. We would like to emphasize that the Bayesian approach used herein to compute upper confidence bounds and the associated proof methodologies can also be adapted to other UCB based approaches that are widely used in online learning (Efroni et al., 2020; Rosenberg and Mansour, 2019) and dynamics estimation (Tarbouriech and Lazaric, 2019).

Related Work Our work is closely related to (Shekhar et al., 2020). We adopt their sampling strategy and many of the proof methodologies. The key difference between our work and (Shekhar et al., 2020) is our Bayesian approach for computing the upper confidence bounds on the variances. Our approach for computing these bounds has a significant impact on the performance in practice, especially in the non-asymptotic regime. Multiple distance metrics have been considered in (Shekhar et al., 2020), whereas we restrict our attention to the mean squared error. It is however possible to extend our methodology to the other distance metrics as well. When the pmfs of interest have a binary support, which is the case in the problem of seroprevalence estimation, then the problem of estimating distributions reduces to the problem of estimating means of random variables. Substantial amount of work has been done in the area of estimating the means of random variables (Carpentier et al., 2011). In (Carpentier et al., 2011), two approaches for computing upper confidence bounds were proposed, one using the Chernoff-Hoeffding bound and the other using the empirical Bernstein bound. The Chernoff-Hoeffding bound approach is very close to that of (Shekhar et al., 2020). We observe that the performance of our Bayesian approach is better than both the approaches in (Carpentier et al., 2011). However, the regret bounds in (Carpentier et al., 2011) for the Bernstein approach are tighter than ours. Our work is similar in spirit to works on multi-armed bandits (Kaufmann et al., 2012; Agrawal and Goyal, 2013) wherein Bayesian (UCB and Thompson sampling) approaches outperform frequentist approaches.

Notation In general, subscripts denote time indices unless stated otherwise. For a sampling strategy g and a collection of pmfs p , we use $\mathbb{P}_p^g[\cdot]$ and $\mathbb{E}_p^g[\cdot]$ to indicate respectively that the probability

and expectation depend on the choice of g , and that they are conditioned on the model p . We denote the indicator function associated with an event \mathcal{E} with $\mathbb{1}_{\mathcal{E}}$.

2. Problem Formulation

We are interested in estimating K probability mass functions (pmfs) over a finite space $\{1, \dots, L\}$ from their samples. Let the k -th pmf be denoted by $p^{(k)} \doteq (p^{(k,1)}, \dots, p^{(k,L)})$ and let $p \doteq [p^{(1)}, \dots, p^{(K)}]$. For convenience, we will refer to each of these pmfs as an *arm*.

The total number of samples that can be obtained is N . At any given time, we can adaptively choose an arm to obtain a sample. The n -th arm sampled is denoted by U_n . Let the outcome of the n -th sample be denoted by Y_n . We have $Y_n \sim p^{(U_n)}$. Let the history of sampled arms and their corresponding outcomes be denoted by $I_n \doteq \{U_1, Y_1, \dots, U_{n-1}, Y_{n-1}\}$. Let the strategy used to collect samples at time n be g_n . In other words, $U_n = g_n(I_n)$. After collecting n samples, we form an estimate $\hat{p}_n^{(k)}$ of each pmf $p^{(k)}$, which in our case, is the empirical distribution. The mean squared error between the true distribution $p^{(k)}$ and the estimate is denoted by

$$\mathcal{L}_N^{(k)}(p^{(k)}, g) = \mathbb{E}_p^g \left[\text{MSE} \left(p^{(k)}, \hat{p}_N^{(k)} \right) \right]. \quad (1)$$

Remark 1 *Note that we can scale the loss $\mathcal{L}^{(k)}$ by an appropriate weight $w^{(k)}$. Our sampling strategy and proofs can be easily adapted to this weighted loss.*

As discussed earlier, our sampling strategy will involve the computation of upper confidence bounds on certain parameters. When Bayesian methods are used to derive upper confidence bounds, it is generally not possible to provide performance guarantees for every instance p of the underlying pmfs. In such cases, the loss \mathcal{L} is *locally averaged* over a small region containing the true pmf p . We will therefore consider a loss that is locally averaged with respect to p using a distribution ϱ . We refer to (Brown et al., 2001; Bayarri and Berger, 2004) for a detailed discussion on the interpretation of local averaging.

Remark 2 *Note that the distribution ϱ used for local averaging is not the same as the prior on p . We will use a Bayesian interpretation of the pmfs p only to compute tighter upper confidence bounds. However, our formulation and performance guarantees are frequentist in nature since the distribution ϱ is supported on an arbitrarily small region around p .*

Let Δ denote the $L - 1$ dimensional simplex and let $\varrho = \varrho^{(1)} \times \dots \times \varrho^{(K)}$ be a product distribution on Δ^K . Our goal is to minimize the following cost associated with a sampling strategy g which is given by

$$J_N(g, \varrho) \doteq \max_k \left\{ \mathbb{E}^{\varrho} \left[\mathcal{L}_N^{(k)}(\pi^{(k)}, g) \right] \right\}, \quad (2)$$

where $\pi \sim \varrho$. The distribution $\varrho^{(k)}$ has the following form. For any Borel measurable set $\mathcal{A} \subseteq \Delta$,

$$\varrho^{(k)}(\mathcal{A}) = \mu \left(\mathcal{A} \cap \eta(p^{(k)}) \right) / \eta, \quad (3)$$

where μ is the uniform distribution over Δ and $\eta(p^{(k)})$ is an ℓ_2 -ball around $p^{(k)}$, such that $\mu(\eta(p^{(k)})) = \eta$. Clearly, the cost in (2) approaches the cost in (Shekhar et al., 2020; Carpentier et al., 2011) when

$\eta \rightarrow 0$ and thus, it is desirable to select small η . However, the performance guarantees we provide in this paper become weaker as η becomes small. We will discuss the effects of the choice of η in subsequent sections.

The Oracle Let there be an oracle that knows the probability distributions p . Using this information, the oracle allocates a fixed number of samples for each arm. If the number of samples allocated to arm k is $T^{(k)}$, the mean-squared error associated with arm k 's estimate is given by

$$\mathbb{E}_p^g \left[\text{MSE} \left(p^{(k)}, \hat{p}_N^{(k)} \right) \right] = \frac{\sum_l p^{(k,l)} (1 - p^{(k,l)})}{T^{(k)}} \quad (4)$$

$$\doteq \frac{c^{(k)}}{T^{(k)}} \doteq \varphi(c^{(k)}, T^{(k)}), \quad (5)$$

where φ is referred to as the *tracking function* and $c^{(k)}$ is referred to as the *tracking parameter* (Shekhar et al., 2020). The oracle solves the following optimization problem to obtain an allocation.

$$\min_{T^{(1)}, \dots, T^{(K)}} \max_k \varphi(c^{(k)}, T^{(k)}) \quad \text{s.t.} \quad \sum_k T^{(k)} = N. \quad (\text{P1})$$

While Problem (P1) is a combinatorial optimization problem, an approximate solution can easily be obtained by making a convex relaxation. The non-integer solution and the optimal value for this problem can be expressed in closed form as

$$T_*^{(k)} = \frac{c^{(k)} N}{\sum_i c^{(i)}}; \quad \varphi^*(p, N) = \frac{\sum c^{(k)}}{N}. \quad (6)$$

The regret with respect to the oracle that knows the underlying distributions is

$$\mathcal{R}_N(g, \varrho) \doteq J_N(g, \varrho) - \mathbb{E}^\varrho [\varphi^*(p, N)]. \quad (7)$$

3. Definitions and Framework

We will first define some important quantities and compute an upper confidence bound on the tracking parameter $c^{(k)}$. This bound will be used in our sampling strategy described in Section 4.

3.1. Prior Belief

Let us assume that each of the pmfs $p^{(k)}$ is independently drawn from a Dirichlet distribution (Kotz et al., 2004) with parameters $\alpha_1^{(k)} \doteq (\alpha_1^{(k,1)}, \dots, \alpha_1^{(k,L)})$. In other words,

$$(p^{(1)}, \dots, p^{(K)}) \sim \mathcal{D}(\alpha_1^{(1)}) \times \dots \times \mathcal{D}(\alpha_1^{(K)}). \quad (8)$$

This prior distribution¹ will be denoted by ρ_1 . We will use the uniform prior which is equivalent to setting $\alpha_1^{(k,l)} = 1$ for every k, l . Thus, ρ_1 is the same as the distribution μ used in (3).

1. Note that this is distinct from the distribution ϱ that was used for local averaging.

Remark 3 Note that the prior ρ_1 is just something we assume. In a frequentist setting with local averaging, we are not provided with any prior information on the pmfs p . Since the distribution ϱ is defined using the uniform distribution μ in (3), the uniform prior $\rho_1 = \mu$ happens to be a convenient choice for computing our confidence bounds and subsequently analyzing them. In some cases however, we do have some prior information on p . A discussion on how to incorporate certain kinds of priors is provided in Appendix D of (Kartik et al., 2020).

3.2. Posterior Update

When the prior on the pmfs is a factored Dirichlet distribution as in (8), the posterior distribution on the pmfs after collecting n samples is $\prod_k \mathcal{D}(\alpha_{n+1}^{(k)})$ (Kotz et al., 2004), where

$$\alpha_{n+1}^{(k,l)} = \alpha_n^{(k,l)} + \mathbb{1}_{(U_n=k)} \mathbb{1}_{(Y_n=l)}. \quad (9)$$

Further, since the pmf $p^{(k)}$ is Dirichlet distributed, the probability $p^{(k,l)}$ is distributed according to the Beta distribution. More precisely, conditioned on information I_n , we have

$$p^{(k,l)} \sim \text{Beta} \left(\alpha_n^{(k,l)}, \alpha_n^{(k,0)} - \alpha_n^{(k,l)} \right), \quad (10)$$

where $\alpha_n^{(k,0)} \doteq \sum_{l=1}^L \alpha_n^{(k,l)}$.

3.3. Upper Confidence Bounds

Let the number of times arm k has been sampled until time n be

$$T_n^{(k)} = \sum_{\tau=1}^{n-1} U_\tau^{(k)}. \quad (11)$$

Based on the posterior distribution of $p^{(k,l)}$, we can compute a lower bound $a_n^{(k,l)}$ and an upper bound $b_n^{(k,l)}$ on $p^{(k,l)}$ such that

$$\mathbb{P}^{\rho_1} [p^{(k,l)} > b_n^{(k,l)} \mid I_n] \leq \delta_n/2 \quad (12)$$

$$\mathbb{P}^{\rho_1} [p^{(k,l)} < a_n^{(k,l)} \mid I_n] \leq \delta_n/2, \quad (13)$$

where $\delta_n = \frac{\delta}{KLn(1+\log N)}$ and $\delta = \eta N^{-5/2}$. Let us define an interval $E_n^{(k,l)} \doteq [a_n^{(k,l)}, b_n^{(k,l)}] \cap E_{n-1}^{(k,l)}$.

This ensures that the interval $E_n^{(k,l)}$ is non-increasing in n . Note that we can compute this interval based on the inverse cdf of the posterior distribution, which is a Beta distribution.

For arm k , let $q_n^{(k)}$ and $u_n^{(k)}$ respectively be the optimal solution and the optimum value of the following quadratic program

$$\max_{q \in \Delta} \sum_l q_l (1 - q_l) \quad \text{s.t. } q_l \in E_n^{(k,l)} \forall l. \quad (14)$$

Notice that since $E_n^{(k,l)}$ is non-increasing in n , the upper bound $u_n^{(k)}$ is also non-increasing in n . Let us define the event \mathcal{E} as

$$\mathcal{E} \doteq \left\{ p^{(k,l)} \in E_n^{(k,l)} \forall k, l, n \right\}. \quad (15)$$

Note that because of the way the upper bound $u_n^{(k,l)}$ is defined in (14), we have $p^{(k,l)}(1 - p^{(k,l)}) \leq u_n^{(k,l)}$ under the event \mathcal{E} . Further, we have $\mathbb{P}^{\rho_1} [\mathcal{E}] \geq 1 - \delta$ because of the union bound $\sum_n \delta_n \leq \delta$.

4. Strategies and Performance Bounds

The main idea is simply to sample the arm that has maximum $\varphi(u_n^{(k)}, T_n^{(k)})$ at time n . The precise strategy is stated as Algorithm 1.

Algorithm 1: Bayesian UCB

```

for  $n = 1$  to  $N$  do
  for  $k = 1$  to  $K$  do
    | Compute  $u_n^{(k)}$ 
  end
  Assign  $U_n = \arg \max_k \varphi(u_n^{(k)}, T_n^{(k)})$ 
  Sample arm  $U_n$  to obtain  $Y_n$  and update the posterior according to (9)
end
return  $\hat{p}_N^{(k)}$ 

```

4.1. Regret Bounds

Using Algorithm 1, we can achieve nearly the same mean squared error as the oracle would have achieved with the knowledge of underlying distributions. The gap between the oracle and Algorithm 1 is characterized by the regret bound below.

Theorem 4 *Using the sampling strategy in Algorithm 1, we have*

$$\mathcal{R}_N(g, \varrho) \leq \mathcal{O} \left(\frac{\ln(\eta)}{N^{\frac{3}{2}}} \right). \quad (16)$$

Our methodology for deriving the regret bound in Theorem 4 is adapted from that in (Shekhar et al., 2020). There are two key distinctions from the approach in (Shekhar et al., 2020): (i) the gap $e_n^{(k)}$ between the upper bound $u_n^{(k)}$ and the true tracking parameter $c^{(k)}$, and (ii) the use of local averaging with respect to the distribution ϱ . These distinctions arise due to the Bayesian nature of the confidence bounds in (14). In (Shekhar et al., 2020), the gap $e_n^{(k)}$ is bounded by a simple closed form expression (see Lemma 3 in (Shekhar et al., 2020)) which is then used to derive the regret bound. Characterizing this gap in our Bayesian setting is non-trivial due to the relatively complicated construction of the upper bound in (14). In the following two lemmas, we first derive a bound on $e_n^{(k)}$.

Lemma 5 *The length of the interval $E_n^{(k,l)}$ satisfies*

$$\text{len} \left(E_n^{(k,l)} \right) \leq \sqrt{\frac{2 \ln \frac{2}{\delta_n}}{(\alpha_n^{(k,0)} + 1)}} \leq \sqrt{\frac{2 \ln \frac{2}{\delta_n}}{(T_n^{(k)} + 1)}}. \quad (17)$$

Proof See Appendix A. ■

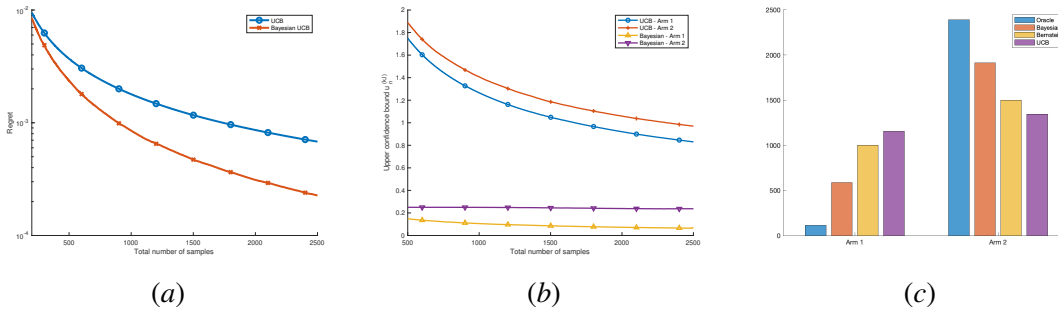


Figure 1: The plot on the left depicts the regret \mathcal{R}_N for the strategy in Algorithm 1 vs. the strategy in (Shekhar et al., 2020). The plot in the middle depicts the upper confidence bounds $u_n^{(k)}$. The plot on the right depicts the expected number of samples obtained and includes the empirical Bernstein approach in (Carpentier et al., 2011) as well.

Lemma 6 Under event \mathcal{E} , we have

$$e_n^{(k)} \doteq u_n^{(k)} - c^{(k)} \leq \sqrt{\frac{8 \ln \frac{2}{\delta_n}}{(T_n^{(k)} + 1)}}. \quad (18)$$

Proof See Appendix B. ■

Note that the bound here is as tight as the one in Lemma 3 of (Shekhar et al., 2020). These lemmas address the first aforementioned distinction from (Shekhar et al., 2020). Using Lemma 6, we can follow the steps of the proof of Theorem 1 in (Shekhar et al., 2020) to complete the proof of Theorem 4. We refer the reader to Appendix C for the full proof with details related to local averaging.

5. Numerical Results

We consider an experimental setup with two arms ($K = 2$) and a binary support ($L = 2$). The pmfs associated with arms 1 and 2 are $[0.99, 0.01]$ and $[0.7, 0.3]$. The total number of samples $N = 2500$ and the value of $\eta = 1/N$. Figure 1(a)subfigure depicts the regret \mathcal{R}_N for the sampling strategy in (Shekhar et al., 2020) and our sampling strategy in Algorithm 1. We will refer to the former strategy as UCB and the latter as Bayesian UCB. There is a clear improvement in performance in terms of regret. We also plot the average number of samples acquired by both strategies in Figure 1(c)subfigure. We observe that the number of samples acquired by our Bayesian UCB approach is closer to the oracle allocation.

Recall that the main difference between the UCB and the Bayesian UCB strategies is the way the upper bound $u_n^{(k,l)}$ is computed. In both these approaches, we observe that for every component, the value of the tracking function $u_{N+1}^{(k)}/T_{N+1}^{(k)}$ is nearly the same for all arms k .

Proposition 7 For every arm k , we have

$$\frac{\sum_k u_{N+1}^{(k)}}{N} \times \frac{1}{f_N} \leq \frac{u_{N+1}^{(k)}}{T_{N+1}^{(k)}} \leq \frac{\sum_k u_{N+1}^{(k)}}{N} \times f_N, \quad (19)$$

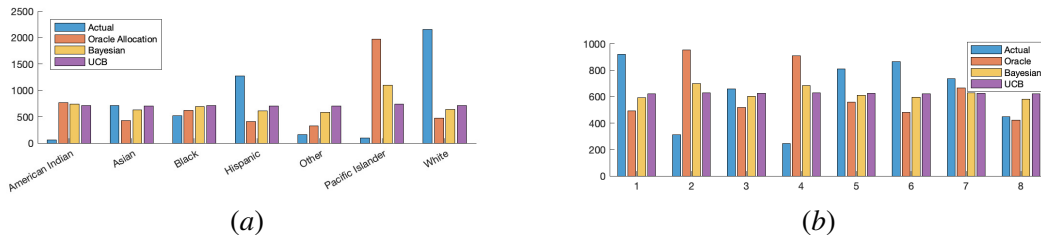


Figure 2: Population classified on the basis of ethnicity and location, respectively in plots 2(a) subfigure and 2(b) subfigure. The plots depict the number of sampled obtained in each category under each strategy. Here, *Actual* refers to the number of samples obtained by the survey (Sood et al., 2020)

where $f_N = 1 + o(1)$. Therefore, $T_{N+1}^{(k)} \propto u_{N+1}^{(k)}$, approximately.

In Figure 1(b) subfigure, we plot a typical realization of the upper bounds $u_n^{(k)}$ as a function of time n . Notice that the bounds computed in our approach are significantly tighter. This tightness can be attributed to two factors: (i) confidence intervals on the probabilities $p^{(k,l)}$ computed using the inverse cdf of the posterior distribution, and (ii) the quadratic program for computing $u_n^{(k)}$ in (14). The bounds computed using the method in (Shekhar et al., 2020) do capture the *difference* between the tracking parameters for arms 1 and 2. However, according to Proposition 7, it is the *ratio* between the upper bounds that matters. The ratio between the upper bounds in (Shekhar et al., 2020) is close to 1 even for $N = 2500$. While this ratio will approach $c^{(1)}/c^{(2)}$ eventually, it is quite slow in practice. While our Bayesian approach provides tighter bounds, there is a computational cost associated with it, especially because of the diminishing probability δ_n . For the binary support case, there are simpler closed form approximations (Bayarri and Berger, 2004; Brown et al., 2001) that can be used to compute $a_n^{(k,l)}$ and $b_n^{(k,l)}$.

We propose to use our sampling strategy for tracking SARS-CoV-2 seroprevalence in populations classified on the basis of location, ethnicity, age etc. Each category of interest can be modeled as an arm with a Bernoulli pmf, where the probability $p^{(k,2)}$ represents the positivity associated with category k . We use data from a randomized seroprevalence survey conducted in Los Angeles county (Sood et al., 2020). The plots in Figures 2(a) subfigure and 2(b) subfigure depict the gap between optimal sample allocation and the random allocation strategy used by the survey. This gap illustrates the importance of adaptive randomized sampling for obtaining uniform estimates of seroprevalence. Without careful sampling, we can have bad estimates for certain groups that may be particularly vulnerable.

Remark 8 For estimating SARS-CoV-2 seroprevalence, there are some practical aspects that require a slightly different formulation and sampling strategy. We provide a heuristic approach for addressing these issues (see Appendix E in (Kartik et al., 2020)) and the analysis of this approach is a problem for future work.

6. Conclusions and Future Work

For the problem of uniformly estimating pmfs in a mean squared sense, a UCB sampling approach is proposed. Unlike the state-of-the-art approach in (Shekhar et al., 2020), a Dirichlet prior is assumed on the pmfs and the upper confidence bounds are computed based on the posterior. It is analytically shown that the performance of this Bayesian approach can be no worse than the approach in (Shekhar et al., 2020), and has noticeably better performance in practice. We discuss the potential application of the sampling approach for SARS-CoV-2 seroprevalence and incidence rate estimation.

There are two potential directions for future work. One is to incorporate temporal variation in the underlying pmfs. This is particularly important in studying SARS-CoV-2 incidence rates. The other is to incorporate a prior on the collection pmfs so that the correlation between them can be exploited to improve our estimates and the sampling strategy. For both these directions, a Bayesian model can be more helpful.

Appendix A. Proof of Lemma 5

A Beta distribution with parameters α, β is sub-Gaussian with parameter $\frac{1}{4(\alpha+\beta+1)}$ (Marchal et al., 2017). And because of (10) we can conclude using the Chernoff bound (Ross, 2014) that

$$\mathbb{P}^{\rho_n} \left[p^{(k,l)} > \frac{\alpha_n^{(k,l)}}{\alpha_n^{(k,0)}} + \sqrt{\frac{\ln \frac{2}{\delta_n}}{2(\alpha_n^{(k,0)} + 1)}} \right] \leq \delta_n \quad (20)$$

$$\mathbb{P}^{\rho_n} \left[p^{(k,l)} < \frac{\alpha_n^{(k,l)}}{\alpha_n^{(k,0)}} - \sqrt{\frac{\ln \frac{2}{\delta_n}}{2(\alpha_n^{(k,0)} + 1)}} \right] \leq \delta_n. \quad (21)$$

Since $a_n^{(k,l)}$ and $b_n^{(k,l)}$ are selected using the inverse cdf of the posterior distribution, we have

$$b_n^{(k,l)} - a_n^{(k,l)} \leq 2 \sqrt{\frac{\ln \frac{2}{\delta_n}}{2(\alpha_n^{(k,0)} + 1)}}. \quad (22)$$

The second inequality in the lemma simply follows from the fact the $\alpha_n^{(k,0)} \geq T_n^{(k)}$.

Appendix B. Proof of Lemma 6

Under event \mathcal{E} , we have

$$|u_n^{(k)} - c^{(k)}| \quad (23)$$

$$= \left| \sum_l \left((p^{(k,l)})^2 - (q_n^{(k,l)})^2 - p^{(k,l)} + q_n^{(k,l)} \right) \right| \quad (24)$$

$$\leq \sum_l^a \left(\left| (p^{(k,l)})^2 - (q_n^{(k,l)})^2 \right| \right) \quad (25)$$

$$= \sum_l \left| p^{(k,l)} - q_n^{(k,l)} \right| \left(p^{(k,l)} + q_n^{(k,l)} \right) \quad (26)$$

$$\leq \sum_l \text{len} \left(E_n^{(k,l)} \right) \left(p^{(k,l)} + q_n^{(k,l)} \right) \quad (27)$$

$$\stackrel{b}{\leq} 2 \sqrt{\frac{\ln \frac{2}{\delta_n}}{2(T_n^{(k)} + 1)}} \sum_l \left(p^{(k,l)} + q_n^{(k,l)} \right) \quad (28)$$

$$= 4 \sqrt{\frac{\ln \frac{2}{\delta_n}}{2(T_n^{(k)} + 1)}}. \quad (29)$$

Inequality in (a) is because of the triangle inequality and the inequality in (b) is due to Lemma 5.

Appendix C. Proof of Theorem 4

We have

$$\delta \geq \mathbb{P}^{\rho_1}[\mathcal{E}^c] = \mathbb{E}^{\rho_1}[\delta_p] \geq \mathbb{E}^{\rho_1}[\delta_p \mathbb{1}_{\eta(p)}(p)] = \mathbb{E}^\varrho[\delta_p] \eta, \quad (30)$$

where $\delta_p \doteq \mathbb{P}^g[\mathcal{E}^c \mid p]$. Note that the last equality in the display above follows from a simple change of measure argument. Since δ_p is the probability of the unfavorable event \mathcal{E}^c conditioned on the model p , we can obtain a regret bound in terms of δ_p without local averaging using the same arguments as in the proofs of Lemma 1 and Theorem 1 in (Shekhar et al., 2020, Appendix D.3). Thus, we have

$$\mathcal{L}_N^{(k)}(p^{(k)}, g) - \varphi^*(p, N) \quad (31)$$

$$\leq \frac{(K+5)LM}{(\lambda_{\min})^2 N^{3/2}} + \frac{\delta_p}{(\lambda_{\min})^2 N} \left(1 + 6M\sqrt{N}\right) + \frac{6(K-1)M^2}{(\lambda_{\min})^3 N^2} + L\delta_p, \quad (32)$$

where

$$M \doteq \frac{\lambda_{\max} \sqrt{8 \log(2/\delta_N)}}{\lambda_{\min} C} = \mathcal{O}\left(\sqrt{\log(\eta N)}\right) \quad (33)$$

$$\lambda_{\max} \doteq \min_{k, \pi \in \eta(p)} \frac{c^{(k)}}{\sum_i c^{(i)}}; \quad \lambda_{\max} \doteq \max_{k, \pi \in \eta(p)} \frac{c^{(k)}}{\sum_i c^{(i)}} \quad (34)$$

$$C \doteq \sum_i c^{(i)}, \quad (35)$$

and $c^{(i)} = \sum_l \pi^{(i,l)}(1 - \pi^{(i,l)})$. Using (30), we can then locally average the regret as

$$\mathbb{E}^\varrho \left[\mathcal{L}_N^{(k)}(\pi^{(k)}, g) - \varphi^*(p, N) \right] \quad (36)$$

$$\leq \frac{(K+5)LM}{(\lambda_{\min})^2 N^{3/2}} + \frac{\delta}{(\eta \lambda_{\min})^2 N} \left(1 + 6M\sqrt{N}\right) + \frac{6(K-1)M^2}{(\lambda_{\min})^3 N^2} + L \frac{\delta}{\eta} \quad (37)$$

$$= \mathcal{O}(\log(\eta) N^{-3/2}). \quad (38)$$

Acknowledgments

This research was supported, in part, by grants ONR N00014-15-1-2550, 503400-78050, NSF CCF-1817200, CCF-2008927, ARO W911NF1910269, Cisco Foundation 1980393, DOE DE-SC0021417 and Swedish Research Council 2018-04359.

References

- Shipra Agrawal and Navin Goyal. Further optimal regret bounds for Thompson sampling. In *Artificial intelligence and statistics*, pages 99–107. PMLR, 2013.
- M Jesús Bayarri and James O Berger. The interplay of Bayesian and frequentist analysis. *Statistical Science*, pages 58–80, 2004.
- Lawrence D Brown, T Tony Cai, and Anirban DasGupta. Interval estimation for a binomial proportion. *Statistical science*, pages 101–117, 2001.
- Alexandra Carpentier, Alessandro Lazaric, Mohammad Ghavamzadeh, Rémi Munos, and Peter Auer. Upper-confidence-bound algorithms for active learning in multi-armed bandits. In *International Conference on Algorithmic Learning Theory*, pages 189–203. Springer, 2011.
- Yonathan Efroni, Shie Mannor, and Matteo Pirota. Exploration-exploitation in constrained MDPs. *arXiv preprint arXiv:2003.02189*, 2020.
- Dhruva Kartik, Neeraj Sood, Urbashi Mitra, and Tara Javidi. Adaptive sampling for estimating distributions: A Bayesian upper confidence bound approach. *arXiv preprint arXiv:2012.04137*, 2020.
- Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On Bayesian upper confidence bounds for bandit problems. In *Artificial intelligence and statistics*, pages 592–600. PMLR, 2012.
- Samuel Kotz, Narayanaswamy Balakrishnan, and Norman L Johnson. *Continuous multivariate distributions, Volume 1: Models and applications*. John Wiley & Sons, 2004.
- Olivier Marchal, Julyan Arbel, et al. On the sub-Gaussianity of the Beta and Dirichlet distributions. *Electronic Communications in Probability*, 22, 2017.
- Aviv Rosenberg and Yishay Mansour. Online convex optimization in adversarial Markov decision processes. In *International Conference on Machine Learning*, pages 5478–5486, 2019.
- Sheldon M Ross. *Introduction to probability models*. Academic press, 2014.
- Shubhanshu Shekhar, Mohammad Ghavamzadeh, and Tara Javidi. Adaptive sampling for estimating probability distributions. In *International Conference on Machine Learning*, 2020.
- Neeraj Sood, Paul Simon, Peggy Ebner, Daniel Eichner, Jeffrey Reynolds, Eran Bendavid, and Jay Bhattacharya. Seroprevalence of SARS-CoV-2-specific antibodies among adults in Los Angeles County, California, on april 10-11, 2020. *Jama*, 2020.
- Jean Tarbouriech and Alessandro Lazaric. Active exploration in Markov decision processes. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 974–982, 2019.