
Identifying Best Interventions through Online Importance Sampling

Rajat Sen^{*1} Karthikeyan Shanmugam^{*2} Alexandros G. Dimakis¹ Sanjay Shakkottai¹

Abstract

Motivated by applications in computational advertising and systems biology, we consider the problem of identifying the best out of several possible *soft interventions* at a *source node* V in an acyclic causal directed graph, to maximize the expected value of a *target node* Y (located downstream of V). Our setting imposes a fixed total budget for sampling under various interventions, along with cost constraints on different types of interventions. We pose this as a best arm identification bandit problem with K arms where each arm is a soft intervention at V , and leverage the information leakage among the arms to provide the first gap dependent error and simple regret bounds for this problem. Our results are a significant improvement over the traditional best arm identification results. We empirically show that our algorithms outperform the state of the art in the Flow Cytometry data-set, and also apply our algorithm for model interpretation of the Inception-v3 deep net that classifies images.

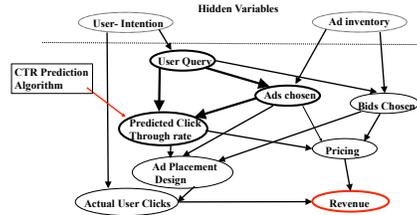


Figure 1. Computational advertising example borrowed from (Bottou et al., 2013). Various observable and hidden variables are shown. The topology of the causal graph is known; however the strengths of most interactions are unknown. A click-rate scoring algorithm predicts future user click through rates from users’ search queries and the set of ads relevant to the user query chosen from an ad inventory. The algorithm’s output determines the ads displayed (as well as the display style), and through a complex causal graph, finally determines actual revenue. The part of the network in bold – distribution of user queries and matching ad keywords is known (including strengths), and the input output characteristics of several candidate (randomized) click-rate scoring algorithms are known. The objective is to choose the best algorithm that maximizes the revenue (target in bold red).

1. Introduction

Causal graphs (Pearl, 2009b) are useful for representing causal relationships among interacting variables in large systems (Bottou et al., 2013). Over the last few decades, causal models have found use in computational advertising (Bottou et al., 2013), biological systems (Meinshausen et al., 2016), sociology (Blalock, 1985), agriculture (Splawa-Neyman et al., 1990) and epidemiology (Joffe et al., 2012). There are two important questions commonly studied with causal graphs: (i) How to learn a directed causal graph that encodes the pattern of interaction among components in a system (*casual structure learning*)? (Pearl, 2009b), and (ii) Using previously acquired (partial) knowledge about the causal graph structure, how to estimate

and/or to optimize the effect of a new *intervention* on other variables (*optimization*) (Bottou et al., 2013; Joffe et al., 2012; Kemmeren et al., 2014; Bonneau et al., 2007; Krouk et al., 2013)? Here, an *intervention* is a forcible change to the value of a variable in a system. The change either alters the relationship between the parental causes and the variable, or decouples it from the parental causes entirely. Our focus is on optimizing over a given set of interventions.

An illustrative example includes online advertising (Bottou et al., 2013), where there is a collection of click-through rate scoring algorithms that provide an estimate of the probability that a user clicks on an ad displayed at a specific position. The interventions occur through the choice of click-through rate scoring algorithm; the algorithm choice directly impacts ad placement and pricing, and through a complex network of interactions, affects the revenue generated through advertisements. The revenue is used to determine the best scoring algorithm (optimize for the best intervention); see Figure 1. Another example is in biological gene-regulatory networks (Bonneau et al., 2007), where a large number of genomes interact amongst each other and

^{*}Equal contribution ¹The University of Texas at Austin ²IBM Thomas J. Watson Research Center. Correspondence to: Rajat Sen <rajat.sen@utexas.edu>.

also interact with environmental factors. The objective here is to understand the best perturbation of some genomes in terms of its effect on the expression of another subset of genomes (target) in cellular systems.

This paper focuses on the following setting: We are given apriori knowledge about the structure and strength of interactions over a small part of the causal graph. In addition, there is freedom to intervene (from a set of allowable interventions) at a certain node in the known part of the graph, and collect data under the chosen intervention; further we can alter the interventions over time and observe the corresponding effects. Given a set of potential interventions to optimize over, the key question of interest is: How to choose the best sequence of T allowable interventions in order to discover which intervention maximizes the expectation of a downstream target node?

Determining the best intervention in the above setting can be cast as a best arm identification bandit problem, as noted in (Lattimore et al., 2016). The possible interventions to optimize over are the arms of the bandit, while the sample value of the *target* node under an intervention is the *reward*.

More formally, suppose that V is a node in a causal graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ (as shown in Fig. 2), with the parents of V denoted by $pa(V)$. In Fig. 1, V corresponds to the *click-through rate* and its parents are *user-query* and *ads-chosen*. This essentially means that V is causally determined by a function of $pa(V)$ and some exogenous random noise. This dependence is characterized by the conditional $P(V|pa(V))$ ¹. Then a (soft) intervention mathematically corresponds to changing this conditional probability distribution i.e. probabilistically forcing V to certain states given its parents. In the computational advertising example, the interventions correspond to changing the click through rate scoring algorithm (i.e. $P(\text{click through rate}|\text{ads chosen}, \text{user query})$), whose input-output characteristics are well-studied. Further, suppose that the effect of an intervention is observed at a node Y which is downstream of V in the topological order (w.r.t \mathcal{G}) -refer to Fig. 2. Then, our key question is stated as follows: *Given a collection of interventions $\{P_0(V|pa(V)), \dots, P_{K-1}(V|pa(V))\}$, find the best intervention among these K possibilities that maximizes $E[Y]$ under a fixed budget of T (intervention, observation) pairs.*

1.1. Main Contributions

(Successive Rejects Algorithm) We provide an efficient *successive rejects* multi-phase algorithm. The algorithm uses *clipped importance sampling*. The clipper level is set *adaptively* in each phase in order to trade-off *bias and vari-*

¹Formally if node V has parents V_1, V_2 , then this distribution is the conditional $P(V = v|V_1 = v_1, V_2 = v_2)$ for all v, v_1, v_2 .

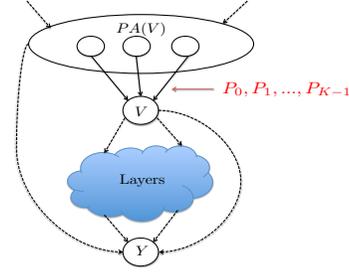


Figure 2. Illustration of our setting. The soft interventions modify $P(V|pa(V))$. The various conditionals and the marginal of $pa(V)$ is assumed to be known from prior data. The *target* variable Y lies further downstream in the unknown portions of the causal graph.

ance. Our procedure yields a major improvement over the algorithm in (Lattimore et al., 2016) (both in theoretical guarantees and in practice), which sets the clipper and allocates samples in a static manner.

(Gap Dependent Error Guarantees under Budget Constraints) In the classic best arm identification problem (Audibert & Bubeck, 2010), Audibert et al. derive *gap* dependent bounds on the probability of error given a fixed sample budget. Specifically, let $\Delta_{(i)}$ be the i -th largest gap (difference) in the expected reward from that of the best arm (e.g. $\Delta_{(1)}$ is the difference between the best arm expected reward and the second best reward). Then, it has been shown in (Audibert & Bubeck, 2010) that the number of samples needed scales as (upto poly log factors) $\max_i(i/\Delta_{(i)}^2)$.

In our setting, a fundamental difference from the classical best arm setting (Audibert & Bubeck, 2010) is the *information leakage* across the arms, i.e. samples from one arm can inform us about the expected value of other arms because of the shared causal graph. We show that this information leakage yields significant improvement both in theory and practice. We derive the first *gap dependent* (gaps between the expected reward at the target under different interventions) bounds on the probability of error in terms of the number of samples T , cost budget B on the relative fraction of times various arms are sampled and the divergences between soft intervention distributions of the arms.

In our result (upto poly log factors) the factor i is replaced by the 'effective variance' of the estimator for arm (i) , i.e. we obtain (with informal notation) $\max_i \sigma_i^2 / \Delta_{(i)}^2$. σ_i can be much smaller than \sqrt{i} (the corresponding term in the results of (Audibert & Bubeck, 2010)). Our theoretical guarantees quantify the improvement obtained by leveraging information leakage, which has been empirically observed in (Bottou et al., 2013). We discuss in more detail in Sections 3.3, about how these guarantees can be exponentially better than the classical ones. We also derive gap dependent simple regret bounds (Section 3.1).

(Novel f -divergence measure for analyzing Importance Sampling) We provide a novel analysis of *clipped importance sampling* estimators, where pairwise f -divergences between the distributions $\{P_k(V|pa(V))\}$, for a carefully chosen function $f(\cdot)$ (see Section E.1) act as the ‘effective variance’ term in the analysis for the estimators (similar to Bernstein’s bound (Bennett, 1962)).

(Extensive Empirical Validation) We demonstrate that our algorithm outperforms the prior works (Lattimore et al., 2016; Audibert & Bubeck, 2010) on the Flow Cytometry data-set (Sachs et al., 2005) (in Section 4.1). We exhibit an innovative application of our algorithm for *model interpretation* of the Inception Deep Network (Szegedy et al., 2015) for image classification (refer to Section 4.2).

Remark 1. *The techniques in this paper can be directly applied to more general settings like (i) the intervention source (V) can be a collection of nodes (\mathcal{V}) and the changes affect the distribution $P(\mathcal{V}|pa(\mathcal{V}))$, where $pa(\mathcal{V})$ is the union of all the parents; (ii) the importance sampling can be applied at a directed cut separating the sources and the targets, provided the effect of the interventions, on the nodes forming the cut can be estimated. Moreover, our techniques can be applied without the complete knowledge of the source distributions. We explain the variations in more detail in Section B in the appendix.*

1.2. Related Work

The problem lies at the intersection of causal inference and best arm identification in bandits. There have been many studies on the classical *best arm identification* in the bandit literature, both in the fixed confidence regime (Kaufmann et al., 2015; Gabillon et al., 2012) and in the fixed budget setting (Audibert & Bubeck, 2010; Chen & Li, 2015; Jamieson et al., 2014; Carpentier & Locatelli, 2016). It was shown recently in (Carpentier & Locatelli, 2016) that the results of (Audibert & Bubeck, 2010) are optimal. The key difference from our work is that, in these models, there is no information leakage among the arms.

There has been a lot of work (Mooij et al., 2016; Hyttinen et al., 2013; Eberhardt, 2008; Hauser & Bühlmann, 2012; Spirtes et al., 2001; Pearl, 2009a; Loh & Bühlmann, 2014; Shanmugam et al., 2015) on *learning casual models* from data and/or experiments and using it to *estimate causal strength* questions of the *counterfactual* nature. One notable work that partially inspired our work is (Bottou et al., 2013) where the causal graph underlying a computational advertising system (like in Bing, Google etc.) is known and the primary interest is to find out how a change in the system would affect some other variable.

At the intersection of *causality and bandits*, (Lattimore et al., 2016) is perhaps most relevant to our setting. It studies the problem of identifying the best hard interventions

on multiple variables (among many), provided the distribution of the parents of the target is known under those interventions. Simple regret bound of order $O(1/\sqrt{T})$ was derived. We assume soft interventions that affect the mechanism between a ‘source’ node and its parents, far away from the target (similar to the case of computational advertising considered in (Bottou et al., 2013)). Further, we derive the first *gap dependent* bounds (that can be exponentially small in T), generalizing the results of (Audibert & Bubeck, 2010). Our formulation can handle general budget constraints on the bandit arms and also recover the problem independent bounds of (Lattimore et al., 2016) (orderwise). Budget constraints in bandit settings have been explored before in (Abernethy et al., 2015; Slivkins, 2013).

In the context of machine learning, *importance sampling* has been mostly used to recondition input data to adhere to conditions imposed by learning algorithms (Sugiyama et al., 2007; Li et al., 2011; Zhang & Zhao, 2014).

2. Problem Setting

In this work, we consider the problem of identifying the best soft intervention, i.e. the one that maximizes the expected value of a certain target variable. The problem setting is best illustrated in Figure 2. Consider a causal graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ that specifies directed causal relationships between the variables \mathcal{V} . We provide a short introduction to causal graphs in Section A (in the appendix) for the sake of completeness. Let Y be a *target* random variable which is downstream in the graph \mathcal{G} ; the expected value of this target variable is the quantity of interest. Consider another random variable V along with its parents $pa(V)$. We assume that there are K possible soft interventions. Each soft intervention is a distinct conditional distribution that dictates the relationship $pa(V) \rightarrow V$. During a soft intervention $k \in [K]$ ($[K] = \{0, 1, \dots, K - 1\}$), the conditional distribution of V given its parents is set to $P_k(V|pa(V))$ and all other relationships in the causal graph are unchanged.

It is assumed that the conditional distributions $P_k(V|pa(V))$ and marginals for $pa(V)$ for $k \in [K]$ are known from past experiments or existing domain knowledge. We only observe samples of Y, V and $pa(V)$, while the rest of the variables in the causal graph may be unobserved under different interventions. For simplicity we assume that the variables $V, pa(V)$ are discrete while the target variable Y may be continuous/discrete and has bounded support in $[0, 1]$. Further, we assume that the various conditionals, i.e. $P_k(V|pa(V))$ are *absolutely continuous* with respect to each other. In the case of discrete distributions, the non-zero supports of these distributions are identical. However, our algorithm can be easily generalized for continuous distributions on V and $pa(V)$ (as in our experiments in Section 4.1). In this setting, we

are interested in the following natural question: *Which of the K soft interventions yield the highest expected value of the target $(\mathbb{E}[Y])$ and what is the misidentification error that can be achieved with a finite total budget T for samples ?*

Remark 2. *Although we may know apriori the joint distribution of $pa(V)$ and V under different interventions, how the change affects another variable Y in the causal graph is unknown and must be learnt from samples. The task is to **transfer prior knowledge** to identify the best intervention.*

Bandit Setting: The K different soft interventions can be thought of as the K arms of a bandit problem. Let the reward of arm k be denoted by: $\mu_k = \mathbb{E}_k[Y]$, where $\mathbb{E}_k[Y]$ is the expected value of Y under the soft intervention when the conditional distribution of V given its parents $pa(V)$ is set to $P_k(V|pa(V))$ (soft intervention k) while keeping all other things in \mathcal{G} unchanged. We assume that there is only one best arm. Let k^* be the arm that yields the highest expected reward and μ^* be the value of the corresponding expected reward i.e. $k^* = \arg \max_k \mu_k$ and $\mu^* = \mu_{k^*}$. Let the optimality gap of the k^{th} arm be defined as $\Delta_k = \mu^* - \mu_k$. We shall see that these gaps $\{\Delta_k\}_{k=0}^{K-1}$ and the relationship between distributions $\{P_k(V|pa(V))\}_{k=0}^{K-1}$ are important parameters in the problem. Let $\Delta = \min_{k \neq k^*} \Delta_k$.

Fixed Budget for Samples: In this paper, we work under the fixed budget setting of best arm identification (Audibert & Bubeck, 2010). Let T_k be the number of times the k^{th} intervention is used to obtain samples. We require that $\sum_{k=0}^K T_k = T$. Let $\nu_k = \frac{T_k}{T}$ be the fraction of times the k^{th} intervention is played.

Additional Cost Budget on Interventions: In the context of causal discovery, some *interventions* require a lot more resources or experimental effort than the others. We find such examples in the context of online advertisement design (Bottou et al., 2013). Therefore, we introduce two variants of an additional cost constraint that influences the choice of interventions. (i) *Difficult arm budget (S1)*: Some arms are deemed to be *difficult*. Let $\mathcal{B} \subset [K]$ be the set of *difficult* arms. We require that the total fraction of times the difficult arms are played does not exceed B i.e. $\sum_{k \in \mathcal{B}} \nu_k \leq B$. (ii) *Cost Budget (S2)*: This is the most general budget setting that captures the variable costs of sampling each arm (Slivkins, 2013). We assume that there is a cost c_k associated with sampling arm k . It is required that the average cost of sampling does not exceed a cost budget B i.e. $\sum_{k=0}^{K-1} c_k \nu_k \leq B$. $\mathbf{c} = [c_1, \dots, c_k]$ along with the total budget T completely defines this budget setting. It should be noted that **S1** is a special case of **S2**.

We note that unless otherwise stated, we work with the most general setting in **S2**. We state some of our results

in the setting **S1** for clearer exposition.

Objectives: There are two main quantities of interest:

(*Probability of Error*): This is the probability of failing to identify the best soft intervention (arm). Let $\hat{k}(T, B)$ be the arm that is predicted to be the best arm at the end of the experiment. Then the probability of error $e(T, B)$ (Audibert & Bubeck, 2010; Carpentier & Locatelli, 2016) is given by,

$$e(T, B) = \mathbb{P}(\hat{k}(T, B) \neq k^*)$$

(*Simple Regret*): Another important quantity that has been analyzed in the best arm identification setting is the simple regret (Lattimore et al., 2016). The simple regret is given by $r(T, B) = \sum_{k \neq k^*} \Delta_k \mathbb{P}(\hat{k}(T, B) = k)$.

3. Our Main Results

In this section we provide our main theoretical contributions. In Section 3.2, we provide a successive rejects style algorithm that leverages the information leakage via importance sampling. Then, we provide theoretical guarantees on the probability of mis-identification ($e(T, B)$) and simple regret ($r(T, B)$) for our algorithm in Section 3.3. In order to explain our algorithm and our results formally, we first describe several key ideas in our algorithm and introduce important definitions in Section 3.1.

3.1. Definitions

Quantifying Information Leakage: Observe that, $\mu_k = \mathbb{E}_k[Y] = \mathbb{E}_{k'} \left[Y \frac{P_k(V|pa(V))}{P_{k'}(V|pa(V))} \right]$. By weighting samples with the correct ratio of conditional probabilities $P_k(\cdot)$ and $P_{k'}(\cdot)$, it is possible to use samples under intervention k' to estimate the mean under intervention k . This is the basic idea behind *importance sampling* which has been used in similar settings (Lattimore et al., 2016; Bottou et al., 2013). However, the variance of this estimator depends on the ratio of $P_k(\cdot)$ to $P_{k'}(\cdot)$. We use a specific f -divergence measure between $P_k(\cdot)$ and $P_{k'}(\cdot)$ to quantify the variance. This f -divergence measure can be calculated analytically or estimated from empirical data (without access to full distributions) using techniques like that of (Pérez-Cruz, 2008).

Definition 1. *Let $f(\cdot)$ be a non-negative convex function such that $f(1) = 0$. For two joint distributions $p_{X,Y}(x, y)$ and $q_{X,Y}(x, y)$ (and the associated conditionals), the conditional f -divergence $D_f(p_{X|Y} \| q_{X|Y})$ is given by:*

$$D_f(p_{X|Y} \| q_{X|Y}) = \mathbb{E}_{q_{X,Y}} \left[f \left(\frac{p_{X|Y}(X|Y)}{q_{X|Y}(X|Y)} \right) \right].$$

Recall that P_i is the conditional distribution of node V given the state of its parents $pa(V)$. Thus, $D_f(P_i \| P_j)$ is the conditional f -divergence between the conditional distributions P_i and P_j . Now we define some *log-divergences*

that are crucial in our analysis.

Definition 2. (M_{ij} measure) Let $f_1(x) = x \exp(x-1) - 1$. We define the following log-divergence measure: $M_{ij} = 1 + \log(1 + D_{f_1}(P_i \| P_j))$, $\forall i, j \in [K]$.

Aggregating Interventional Data: We describe an efficient estimator of $\mathbb{E}_k[Y]$ ($\forall k \in [K]$) that combines available samples from different arms. This estimator adaptively weights samples depending on the relative M_{ij} measures, and also uses clipping to control variance by introducing bias. The estimator is given by (1).

Suppose we obtain τ_i samples from arm $i \in [K]$. Let the total number of samples from all arms be denoted by τ . Further, let us index all the samples by $s \in \{1, 2, \dots, \tau\}$, and $\mathcal{T}_k \subset \{1, 2, \dots, \tau\}$ be the indices of all the samples collected from arm k . Let $X_j(s)$ denotes the sample collected for random variable X under intervention j , at time instant s . Finally, let $Z_k = \sum_{j \in [K]} \tau_j / M_{kj}$. We denote the estimate of μ_k by \hat{Y}_k^ϵ (ϵ is an indicator of the level of confidence desired). Our estimator is:

$$\hat{Y}_k^\epsilon = \frac{1}{Z_k} \sum_{j=0}^K \sum_{s \in \mathcal{T}_j} \frac{1}{M_{kj}} Y_j(s) \frac{P_k(V_j(s) | pa(V)_j(s))}{P_j(V_j(s) | pa(V)_j(s))} \times \mathbb{1} \left\{ \frac{P_k(V_j(s) | pa(V)_j(s))}{P_j(V_j(s) | pa(V)_j(s))} \leq 2 \log(2/\epsilon) M_{kj} \right\}. \quad (1)$$

In other words, \hat{Y}_k^ϵ is the weighted average of the clipped samples, where the samples from arm j are weighted by $1/M_{kj}$ and clipped at $2 \log(2/\epsilon) M_{kj}$. The **choice of ϵ** controls the *bias-variance tradeoff* which we will adaptively change in our algorithm. This adaptive clipping is crucial in deriving gap dependent bounds².

3.2. Algorithm

Now, we describe our main algorithmic contribution - Algorithm 1 and 2. Algorithm 1 starts by having all the K arms under consideration and then proceeds in phases, possibly rejecting one or more arms at the end of each phase.

At every phase, Estimator (1) with a phase specific choice of the ϵ parameter (i.e. controlling bias variance trade-off), is applied to all arms under consideration. Using a phase specific threshold on these estimates, some arms are rejected at the end of each phase. A random arm among the ones surviving at the end of *all* phases is declared to be the optimal. We now describe the duration of various phases.

Recall the parameters T - Total sample budget available and B - average cost budget constraint. Let $n(T) =$

²We note that the authors in (Lattimore et al., 2016) discuss the possibility of a multi-phase approach, where clipper levels could change across phases. However, they do not pursue this direction (no specific algorithm or results) as their objective is to derive gap independent bounds (minimax regret).

$\lceil \log 2 \times \log 10 \sqrt{T} \rceil$. Let $\overline{\log}(n) = \sum_{i=1}^n (1/i)$. We will have an algorithm with $n(T)$ phases numbered by $\ell = 1, 2, \dots, n(T)$. Let $\tau(\ell)$ be the total number of samples in phase ℓ . We set $\tau(\ell) = T / (\overline{\log}(n(T)))$ for $\ell \in \{1, \dots, n(T)\}$. Note that $\sum_{\ell} \tau(\ell) = T$. Let \mathcal{R} be the set of arms remaining to compete with the optimal arm at the beginning of phase ℓ which is continuously updated.

Let $\tau_k(\ell)$ be the samples allocated to arm k in phase ℓ . Let $\boldsymbol{\tau}(\ell)$ be the vector consisting of entries $\{\tau_k(\ell)\}$. The vector of allocated samples, i.e. $\boldsymbol{\tau}(\ell)$ is decided by Algorithm 3. Intuitively, an arm that provides sufficient information about all the remaining arms needs to be given more budget than other less informative arms. This allocation depends on the average budget constraints and the relative log divergences between the arms (Definition 2). Algorithm 3 formalizes this intuition, and ensures that variance of the worst estimator (of the form (1)) for the arms in \mathcal{R} is as good as possible (quantified in Theorem 4 and Lemma 4).

Algorithm 1 Successive Rejects with Importance Sampling -v1 (SRISv1) - Given total budget T and the cost budget B (along with c_i 's) picks the best arm.

- 1: SRIS($B, \{M_{kj}\}, T$)
 - 2: $\mathcal{R} = [K]$.
 - 3: Form the matrix $\mathbf{A} \in \mathbb{R}^{K \times K}$ such that $A_{kj} = \frac{1}{M_{kj}}$.
 - 4: **for** $\ell = 1$ to $n(T)$ **do**
 - 5: $\boldsymbol{\tau}(\ell) = \text{ALLOCATE}(\mathbf{c}, B, \mathbf{A}, \mathcal{R}, \tau(\ell))$ (Algorithm 3)
 - 6: Use arm k , $\tau_k(\ell)$ times and collect samples $(Y, V, pa(V))$.
 - 7: **for** $k \in \mathcal{R}$ **do**
 - 8: Let \hat{Y}_k be the estimator for arm k as in (1) calculated with $\{M_{kj}\}$, $\epsilon = 2^{-(\ell-1)}$ and the samples obtained in Line 6.
 - 9: **end for**
 - 10: Let $\hat{Y}_H = \arg \max_{k \in \mathcal{R}} \hat{Y}_k$.
 - 11: $\mathcal{R} = \mathcal{R} - \{k \in \mathcal{R} : \hat{Y}_H > \hat{Y}_k + 5/2^\ell\}$.
 - 12: **if** $|\mathcal{R}| = 1$ **then**
 - 13: **return:** the arm in \mathcal{R} .
 - 14: **end if**
 - 15: **end for**
 - 16: **return:** A randomly chosen arm from \mathcal{R} .
-

Remark 3. Note that Line 6 uses only the samples acquired in that phase. Clearly, a natural extension is to modify the algorithm to re-use all the samples acquired prior to that step. We give that variation in Algorithm 2. We prove all our guarantees for Algorithm 1. We conjecture that the second variation has tighter guarantees (dropping a multiplicative log factor) in the sample complexity requirements.

The inverse of the maximal objective of the LP in Algorithm 3 acts as *effective standard deviation* uniformly for

Algorithm 2 Successive Rejects with Importance Sampling -v2 (SRISv2) - Given total budget T and the cost budget B (along with \mathbf{c}) picks the best arm.

- 1: Identical to Algorithm 1 except for Line 6 where all samples acquired in all the phases till that Line is used.

all the estimators for the remaining arms in \mathcal{R} . It is analogous to the variance terms appearing in Bernstein-type concentration bounds (refer to Lemma 4 in the appendix).

Definition 3. The *effective standard deviation* for budget B and arm set $\mathcal{R} \subseteq [K]$ is defined as $\sigma^*(B, \mathcal{R}) = 1/v^*(B, \mathcal{R})$ from Algorithm 3 with input B and arm set \mathcal{R} .

Algorithm 3 Allocate - Allocates a given budget τ among the arms to reduce variance.

- 1: ALLOCATE($\mathbf{c}, B, \mathbf{A}, \mathcal{R}, \tau$)
- 2: Solve the following LP:

$$\begin{aligned} \frac{1}{\sigma^*(B, \mathcal{R})} = v^*(B, \mathcal{R}) &= \max_{\boldsymbol{\nu}} \min_{k \in \mathcal{R}} [\mathbf{A}\boldsymbol{\nu}]_k & (2) \\ \text{s.t. } \sum_{i=0}^K c_i \nu_i &\leq B \text{ and } \sum_{j=0}^K \nu_j = 1, \nu_i \geq 0. \end{aligned}$$

- 3: Assign $\tau_j = \nu_j^*(B, \mathcal{R})\tau$

3.3. Theoretical Guarantees

We state our main results as Theorem 1 and Theorem 2, which provide guarantees on probability of error and simple regret respectively. Our results can be interpreted as a natural generalization of the results in (Audibert & Bubeck, 2010), when there is information leakage among the arms. This is the first gap dependent characterization.

Theorem 1. (Proved formally as Theorem 5) Let $\Delta = \min_{k \neq k^*} \Delta_k$. Let $\sigma^*(\cdot)$ be the effective standard deviation as in Definition 3. The probability of error for Algorithm 1 satisfies:

$$e(T, B) \leq 2K^2 \log_2(20/\Delta) \exp\left(-\frac{T}{2\bar{H}\log(n(T))}\right) \quad (3)$$

when the budget for the total number of samples is T and $\Delta \geq 10/\sqrt{T}$. Here,

$$\bar{H} = \max_{k \neq k^*} \log_2(10/\Delta_k)^3 \left(\frac{\sigma^*(B, \mathcal{R}^*(\Delta_k))}{\Delta_k} \right)^2 \quad (4)$$

and $\mathcal{R}^*(\Delta_k) = \{s : \log_2\left(\frac{10}{\Delta_s}\right) \geq \lfloor \log_2\left(\frac{10}{\Delta_k}\right) \rfloor\}$ is the set of arms whose distance from the optimal arm is roughly at most twice that of arm k .

Comparison with the result in (Audibert & Bubeck, 2010): Let $\tilde{\mathcal{R}}(\Delta_k) = \{s : \Delta_s \leq \Delta_k\}$, i.e. the set of arms which are closer to the optimal than arm k . Let $\tilde{H} = \max_{k \neq k^*} \frac{|\tilde{\mathcal{R}}(\Delta_k)|}{\Delta_k^2}$. The result in (Audibert & Bubeck, 2010) can be stated as: *The error in finding the optimal arm is bounded as: $e(T) \leq O\left(K^2 \exp\left(-\frac{T-K}{\log(K)\tilde{H}}\right)\right)$.*

Our work is analogous to the above result (upto poly log factors) except that \bar{H} appears instead of \tilde{H} . In Section D.1 (in the appendix), we demonstrate through simple examples that $\sigma^*(B, \mathcal{R}^*(\Delta_k))$ can be significantly smaller than $\sqrt{|\tilde{\mathcal{R}}(\Delta_k)|}$ (the corresponding term in $e(T)$ above) even when there are no average budget constraints. Moreover, our results can be exponentially better in the presence of average budget constraints (examples in Section D.1). Now we present our bounds on simple regret in Theorem 2.

Theorem 2. (Proved formally as Theorem 5) Let $\sigma^*(\cdot)$ be the effective standard deviation as in Definition 3. The simple regret of Algorithm 1 when the number of samples is T satisfies:

$$\begin{aligned} r(T, B) &\leq \frac{10}{\sqrt{T}} \mathbb{1}\left\{\exists k \neq k^* \text{ s.t. } \Delta_k < 10/\sqrt{T}\right\} + \\ &2K^2 \sum_{\substack{k \neq k^* \\ \Delta_k \geq 10/\sqrt{T}}} \Delta_k \log_2\left(\frac{20}{\Delta_k}\right) \exp\left(-\frac{T}{2\bar{H}_k \log(n(T))}\right) \end{aligned} \quad (5)$$

Here, $\bar{H}_k = \max_{l: \Delta_l \geq \Delta_k} \frac{\log_2(10/\Delta_l)^3}{(\Delta_l/10)^2 v^*(B, \mathcal{R}^*(\Delta_l))^2}$ and $\mathcal{R}^*(\Delta_k) = \{s : \log_2\left(\frac{10}{\Delta_s}\right) \geq \lfloor \log_2\left(\frac{10}{\Delta_k}\right) \rfloor\}$.

Comparison with the result in (Lattimore et al., 2016): In (Lattimore et al., 2016), the simple regret scales as $O(1/\sqrt{T})$ and does not adapt to the gaps. We provide gap dependent bounds that can be exponentially better than that of (Lattimore et al., 2016) (when Δ_k 's are not too small and the first term in (5) is zero). More over our bounds generalize to gap independent bounds that match $O(1/\sqrt{T})$. Further details are provided in Section D.2 (in the appendix).

We defer the theoretical analysis to Section E. Theorem 1 and Theorem 2 are subparts of our main technical theorem (Theorem 5), which is proved in Section E.5.

4. Empirical Validation

We empirically validate the performance of our algorithms in two real data settings. In Section 4.1, we study the empirical performance of our algorithm on the flow cytometry data-set (Sachs et al., 2005). In Section 4.2, we apply our algorithms for the purpose of *model interpretability* of the Inception Deep Network (Szegedy et al., 2015) in the context of image classification. In Section F (in the appendix)

we include more experiments. In the appendix we empirically show that our divergence metric is fundamental and replacing it with other divergences is sub-optimal.

4.1. Flow Cytometry Data-Set

The flow cytometry data-set (Sachs et al., 2005) (extensively used for validating causal inference algorithms) consists of multi-variate measurements of protein interactions in a *single cell*, under different experimental conditions (*soft interventions*). Our experiments are aimed at *identifying* the best intervention among many, given some ground truth about the causal graph. For, this purpose we borrow the causal graph from Fig. 5(c) in (Mooij & Heskes, 2013) (shown in Fig. 3a) and consider it to be the ground truth.

Parametric linear models have been popularly used for causal inference on this data-set (Meinshausen et al., 2016; Cho et al., 2016). We fit a GLM gamma model (Hardin et al., 2007) between the activation of each node and its parents in Fig. 3a using the observational data. In Section F.2 (in the appendix) we provide further details showing that the sampled distributions in the fitted model are extremely close to the empirical distributions from the data. The soft interventions signifying the arms are generated by changing the distribution of a source node *pkc* in the GLM. The objective is to identify the intervention that yields the highest output at the target node *erk*³ We provide empirical results for two sets of interventions at the source node. Both these experiments have been performed with 15 arms each representing different distributions at *pkc*.

Budget Restriction: The experiments are performed in the budget setting **S1**, where all arms *except* arm 0 are deemed to be *difficult*. We plot our results as a function of the total samples T , while the fractional budget of the *difficult* arms (B) is set to $1/\sqrt{T}$. Therefore, we have $\sum_{k \neq 0} T_k \leq \sqrt{T}$. This essentially belongs to the case when there is a lot of data that can be acquired for a default arm while any new change requires significant cost in acquiring samples.

Competing Algorithms: We test our algorithms on different problem parameters and compare with related prior work (Audibert & Bubeck, 2010; Lattimore et al., 2016). The algorithms compared are (i) *SRISv1*: Algorithm 1 introduced in Section 3.2. The divergences, $D_{f_1}(P_i||P_j)$ are estimated from sampled data using techniques from (Pérez-Cruz, 2008); (ii) *SRISv2*: Algorithm 2 as detailed in Section 3.2; (iii) *SR*: Successive Rejects Algorithm from (Audibert & Bubeck, 2010) adapted to the budget setting. The division of the total budget T into $K - 1$ phases is identical,

³The activations of the node *erk* have been scaled so that the mean is less than one. Note that the marginal distribution still has an exponential tail, and thus does not strictly adhere to our boundedness assumption on the target variable. However, the experiments suggest that our algorithms still perform extremely well.

while the individual arm budgets are decided in each phase according to the budget restrictions; (iv) *CR*: Algorithm 2 from (Lattimore et al., 2016). The optimization problem for calculating the mixture parameter η is not efficiently solvable for general distributions and budget settings. Therefore, the mixture proportions are set by Algorithm 3.

In these experiments, the budget restrictions imply that arm 0 can be pulled much more than the other arms. Intuitively the divergences of the arms from arm 0 as well as the gap Δ defines the hardness of identification. Fig. 3b represents a difficult scenario where the divergences $M_{k0} > 400$ for many arms (large divergences imply low information leakage) and $\Delta = 0.01$ (small Δ increases hardness). In Fig. 3c (easier scenario) the divergences $M_{k0} < 20$ for most arms while the gap is same as before. We see that *SRISv2* outperforms all the other algorithms by a large margin, especially in the low sample regime.

4.2. Interpretability of Inception Deep Network

In this section we use our algorithm for *model interpretation* of the pre-trained Inception-v3 network (Szegedy et al., 2015) for classifying images. *Model Interpretation* essentially addresses: ‘*why does a learning model classify in a certain way?*’, which is an important question for complicated models like deep nets (Ribeiro et al., 2016).

When an RGB image is fed to Inception, it produces an ordered sequence of 1000 labels (e.g ‘drums’, ‘sunglasses’) and generally the top-10 labels are an accurate description of the objects in the image. To address interpretability, we segment the image into a number of *superpixels*/segments (using segmentation algorithms like SLIC (Achanta et al., 2010)) and infer which superpixels encourage the neural net to output a certain label (henceforth referred to as *label-I*; e.g ‘drum’) in top- k (e.g. $k = 10$), and to what extent.

Given a *mixture distribution* over the superpixels of an image (Figure 4a), a few superpixels are randomly sampled from the distribution with replacement. Then a new image is generated where all other superpixels of the original image are blurred out except the ones selected. This image is then fed to Inception, and it is observed whether *label-I* appears within the top- k labels. A *mixture distribution is said to be a good interpretation* for *label-I* if there is a high probability that *label-I* appears for an image generated by this mixture distribution. To empirically test the goodness of a mixture distribution, we would generate (using this mixture distribution) a number of random images, and determine the fraction of images for which *label-I* appears; a large fraction indicates that the mixture distribution is a good interpretation of *label-I*.

Motivated by the above discussion, we generate a large number (3200) of mixture distributions, with the goal of

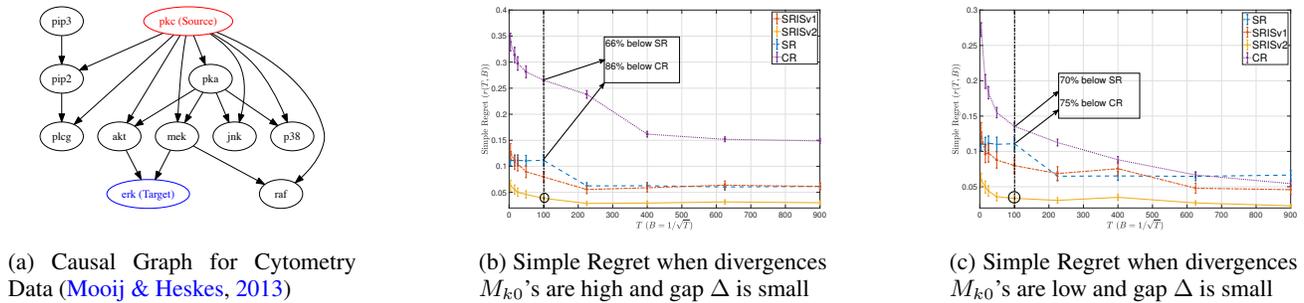


Figure 3. Performance of various algorithms on the cytometry data under different scenarios. The results are averaged over the course of 500 independent experiments. The total sample budget T is plotted on the x -axis. The budget for all arms other than arm 0 is constrained to be less than \sqrt{T} . Here $K = 15$. The performance improvement is especially evident in the low sample regime. For, example in (b) SRISv2 provides more than 65% improvement over CR and SR. It is significant in biological data where number of samples is generally small. SR does not use information leakage while the static clipper in CR cannot adapt to high divergences like in (b).

finding the one that best interprets *label-1*. To highlight the applicability of our algorithm, we allow images to be generated for only 200 of these mixture distributions; in other words, most of the mixture distributions cannot be directly tested. Nevertheless, we determine the best from among the entire collection of mixture distributions (*counterfactuals*).

Specifically in our experiments, we consider the image in Figure 4a, partition it into 43 superpixels, and generate images from mixture distributions by sampling 5 superpixels (with replacement). We generate 3000 arm distributions which lie in the 43-dimensional simplex but have *sparse* support (sparsity of 10 in our examples). The support of these distributions are randomly generated by techniques like markov random walk (encourages contiguity), random choice, etc. as detailed in Section F.3 in the appendix. We are *only* allowed to sample using a different set of 200 arms that are dense distributions chosen uniformly at random from the 43-dimensional simplex. The distributions are generated in a manner which is *completely agnostic* to the image content. The total sample budget (T) is 2500.

Figure 4 shows images in which the segments are weighted in proportion to the optimal distribution (obtained by SRISv2) for the interpretation of three different labels. This showcases the true counterfactual power of the algorithm, as the set of arms that can be optimized over are *disjoint* from the arms that can be sampled from. Moreover the sample budget is *less* than the number of arms. This is an extreme special case of budget setting S2. We see that our algorithm can generate meaningful interpretations for all the labels with relatively less number of runs of Inception. Even sampling 10 times from each of the arms to be optimized over would require 30,000 runs of Inception for a single image and label, while we use only 2500 runs by leveraging information leakage.

Conclusion: In this paper we provide the first gap depen-

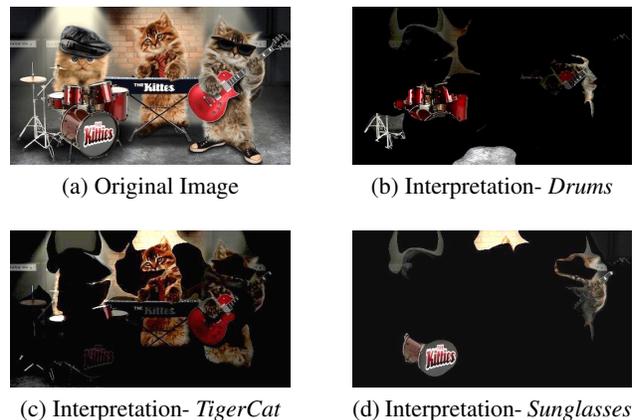


Figure 4. Interpretation for different top labels for the image in (a); Image courtesy (cat). The best mixture distribution generates an image where the highlighted superpixels are most indicative of a label. For example, in (b) we see the drums highlighted, while in (d) (a different mixture distribution), sunglasses are in the focus.

dent error and simple regret bounds for identifying the best soft intervention at a *source* node i.e the one that maximizes the expected value of a downstream *target* node. These bounds are a generalization of the classical best arm identification bounds (Audibert & Bubeck, 2010), when there is information leakage among the arms. We test our algorithm empirically on the flow cytometry data-set and also use it for *interpretability* of Inception-v3 deep net.

Acknowledgements

This work is partially supported by NSF grants CNS-1320175, CCF-1344364, 1407278, 1422549, 1618689, ARO grants W911NF-15-1-0227, W911NF-14-1-0258, W911NF-16-1-0377 and the US DoT supported D-STOP Tier 1 University Transportation Center.

References

- Image source. <http://bit.ly/2hvIwP>. Accessed: 2017-02-23.
- Abernethy, Jacob, Chen, Yiling, Ho, Chien-Ju, and Waggoner, Bo. Low-cost learning via active data procurement. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pp. 619–636. ACM, 2015.
- Achanta, Radhakrishna, Shaji, Appu, Smith, Kevin, Lucchi, Aurelien, Fua, Pascal, and Süsstrunk, Sabine. Slic superpixels. Technical report, 2010.
- Audibert, Jean-Yves and Bubeck, Sébastien. Best arm identification in multi-armed bandits. In *COLT-23th Conference on Learning Theory-2010*, pp. 13–p, 2010.
- Bennett, George. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.
- Blalock, Hubert M. *Causal models in the social sciences*. Transaction Publishers, 1985.
- Bonneau, Richard, Facciotti, Marc T, Reiss, David J, Schmid, Amy K, Pan, Min, Kaur, Amardeep, Thorsson, Vestein, Shannon, Paul, Johnson, Michael H, Bare, J Christopher, et al. A predictive model for transcriptional control of physiology in a free living cell. *Cell*, 131(7):1354–1365, 2007.
- Bottou, Léon, Peters, Jonas, Candela, Joaquin Quinonero, Charles, Denis Xavier, Chickering, Max, Portugaly, Elon, Ray, Dipankar, Simard, Patrice Y, and Snelson, Ed. Counterfactual reasoning and learning systems: the example of computational advertising. *Journal of Machine Learning Research*, 14(1):3207–3260, 2013.
- Carpentier, Alexandra and Locatelli, Andrea. Tight (lower) bounds for the fixed budget best arm identification bandit problem. *arXiv preprint arXiv:1605.09004*, 2016.
- Chen, Lijie and Li, Jian. On the optimal sample complexity for best arm identification. *arXiv preprint arXiv:1511.03774*, 2015.
- Cho, Hyunghoon, Berger, Bonnie, and Peng, Jian. Reconstructing causal biological networks through active learning. *PloS one*, 11(3):e0150611, 2016.
- Eberhardt, Frederick. Almost optimal intervention sets for causal discovery. In *Proceedings of 24th Conference in Uncertainty in Artificial Intelligence (UAI)*, pp. 161–168, 2008.
- Gabillon, Victor, Ghavamzadeh, Mohammad, and Lazaric, Alessandro. Best arm identification: A unified approach to fixed budget and fixed confidence. In *Advances in Neural Information Processing Systems*, pp. 3212–3220, 2012.
- Hardin, James William, Hilbe, Joseph M, and Hilbe, Joseph. *Generalized linear models and extensions*. Stata press, 2007.
- Hauser, Alain and Bühlmann, Peter. Two optimal strategies for active learning of causal networks from interventional data. In *Proceedings of Sixth European Workshop on Probabilistic Graphical Models*, 2012.
- Hyttinen, Antti, Eberhardt, Frederick, and Hoyer, Patrik. Experiment selection for causal discovery. *Journal of Machine Learning Research*, 14:3041–3071, 2013.
- Jamieson, Kevin G, Malloy, Matthew, Nowak, Robert D, and Bubeck, Sébastien. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *COLT*, volume 35, pp. 423–439, 2014.
- Joffe, Michael, Gambhir, Manoj, Chadeau-Hyam, Marc, and Vineis, Paolo. Causal diagrams in systems epidemiology. *Emerging themes in epidemiology*, 9(1):1, 2012.
- Kaufmann, Emilie, Cappé, Olivier, and Garivier, Aurélien. On the complexity of best arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 2015.
- Kemmeren, Patrick, Sameith, Katrin, van de Pasch, Loes AL, Benschop, Joris J, Lenstra, Tineke L, Margaritis, Thanasis, O'Duibhir, Eoghan, Apweiler, Eva, van Wageningen, Sake, Ko, Cheuk W, et al. Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell*, 157(3):740–752, 2014.
- Krouk, Gabriel, Lingeman, Jesse, Colon, Amy Marshall, Coruzzi, Gloria, and Shasha, Dennis. Gene regulatory networks in plants: learning causality from time and perturbation. *Genome biology*, 14(6):1, 2013.
- Lattimore, Finnian, Lattimore, Tor, and Reid, Mark D. Causal bandits: Learning good interventions via causal inference. In *Advances In Neural Information Processing Systems*, pp. 1181–1189, 2016.
- Li, Lihong, Chu, Wei, Langford, John, and Wang, Xuanhui. Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 297–306. ACM, 2011.

- Loh, Po-Ling and Bühlmann, Peter. High-dimensional learning of linear causal networks via inverse covariance estimation. *Journal of Machine Learning Research*, 15 (1):3065–3105, 2014.
- Meinshausen, Nicolai, Hauser, Alain, Mooij, Joris M, Peters, Jonas, Versteeg, Philip, and Bühlmann, Peter. Methods for causal inference from gene perturbation experiments and validation. *Proceedings of the National Academy of Sciences*, 113(27):7361–7368, 2016.
- Mooij, Joris and Heskes, Tom. Cyclic causal discovery from continuous equilibrium data. *arXiv preprint arXiv:1309.6849*, 2013.
- Mooij, Joris M, Peters, Jonas, Janzing, Dominik, Zscheischler, Jakob, and Schölkopf, Bernhard. Distinguishing cause from effect using observational data: methods and benchmarks. *Journal of Machine Learning Research*, 17 (32):1–102, 2016.
- Pearl, Judea. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2009a.
- Pearl, Judea. *Causality*. Cambridge university press, 2009b.
- Pérez-Cruz, Fernando. Kullback-leibler divergence estimation of continuous distributions. In *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pp. 1666–1670. IEEE, 2008.
- Ribeiro, Marco Tulio, Singh, Sameer, and Guestrin, Carlos. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144. ACM, 2016.
- Sachs, Karen, Perez, Omar, Pe'er, Dana, Lauffenburger, Douglas A, and Nolan, Garry P. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.
- Shanmugam, Karthikeyan, Kocaoglu, Murat, Dimakis, Alexandros G, and Vishwanath, Sriram. Learning causal graphs with small interventions. In *Advances in Neural Information Processing Systems*, pp. 3195–3203, 2015.
- Slivkins, Aleksandrs. Dynamic ad allocation: Bandits with budgets. *arXiv preprint arXiv:1306.0155*, 2013.
- Spirtes, Peter, Glymour, Clark, and Scheines, Richard. *Causation, Prediction, and Search*. A Bradford Book, 2001.
- Splawa-Neyman, Jerzy, Dabrowska, DM, Speed, TP, et al. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5(4):465–472, 1990.
- Sugiyama, Masashi, Krauledat, Matthias, and MÄžller, Klaus-Robert. Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, 8(May):985–1005, 2007.
- Szegedy, Christian, Liu, Wei, Jia, Yangqing, Sermanet, Pierre, Reed, Scott, Anguelov, Dragomir, Erhan, Dumitru, Vanhoucke, Vincent, and Rabinovich, Andrew. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015.
- Zhang, Tong and Zhao, Peiling. Stochastic optimization with importance sampling for regularized loss minimization. *arXiv preprint arXiv:1401.2753*, 2014.