Online Zero-Cost Learning: Optimizing Large Scale Network Rare Threats Simulation

Tingwei Liu, Hong Xie, *Member, IEEE*, and John C. S. Lui, *Fellow, IEEE*

Abstract—To provide fast and accurate risk evaluation on network rare threats, importance sampling (IS) is widely used in the rare threat simulation; however, it becomes costly to deal with *many rare threats* simultaneously. For example, a rare threat can be the failure to provide quality-of-service (QoS) guarantees to a critical network flow. Considering network providers often need to deal with many critical flows (i.e., rare threats) simultaneously, if using IS, network providers have to simulate each rare threat with its customized importance distribution individually. To reduce such simulation cost, we propose an efficient mixture importance distribution to simulate multiple rare threats, and then formulate a mixture importance sampling optimization problem (MISO) to select the optimal mixture. We first show that it is challenging to locate the optimal mixture for the "*search direction*" is computationally expensive to evaluate. We then formulate an online learning (OL) framework to estimate the "*search direction*" and learn the optimal mixture from simulation samples of threats. And our OL framework has a "*zero learning cost*" as the samples generated in the learn phase can be reused to provide accurate estimation on the rare threats. We develop two multi-armed bandit OL algorithms so as to: (1) Minimize the sum of estimation variances with a regret of $(\ln T)^2/T$; and (2) Minimize the simulation cost with a regret of $\sqrt{\ln T/T}$, where T denotes the number of simulation samples. We demonstrate the versatility of our method on different network applications. When compared with the uniform mixture IS, our method reduces cost measures (i.e., sum of estimation variances and simulation cost) by as high as 61.6 percent in the Internet backbone network scenario.

Index Terms—Large scale network rare threats simulation, mixture importance sampling, online learning

1 Introduction

In simulating highly fault-tolerant systems like large scale networks or mobile 5G networks, we often have to deal with rare threats; these are events that occur rarely but have catastrophic impacts or consequences. They could potentially arise in many networking applications, and are not easily accessible for the information-bearing signals often lie in a broad set of irrelevant events (i.e., causes). For instance, in communication networks, the network component failures caused by some undesirable events (e.g., equipment ageing or power shortage) can significantly degrade the intended network service [1]. On the Internet, some unexpected node and link failures can result in the undeliveries of critical flows. In smart grids, the network component damages caused by some sudden unforeseen events (e.g., excessive load demand or lighting) may give rise to large-scale blackouts [2]. To quantify such *rare threats*, one needs to evaluate their risks (i.e., occurrence probabilities) accurately. For large and complex networks, this is computationally expensive as illustrated below.

- *Tingwei Liu and John C. S. Lui are with the Department of Computer Science and Engineering, Chinese University of Hong Kong, Hong Kong. E-mail: tingweiliu2013@gmail.com, cslui@cse.cuhk.edu.hk.*
- *Hong Xie is with the Chongqing Key Laboratory of Software Theory and Technology, Chongqing University, Chongqing 400044, China. E-mail: xiehong2018@cqu.edu.cn.*

Manuscript received 12 Sept. 2020; revised 4 Mar. 2021; accepted 18 Apr. 2021. Date of publication 22 Apr. 2021; date of current version 5 Dec. 2022.
(Corresponding author: Hong Xie.)
Digital Object Identifier no. 10.1109/TMC.2021.3074920

Example 1. Consider a rare threat $\mathcal{E}$, which corresponds to the failure of providing promised quality-of-service (QoS) guarantees for a critical flow in a large-scale network. The QoS guarantees for the critical flow is influenced by the status (i.e., operational or failed) of $M\in\mathbb{N}_+$ components (i.e., links or nodes) of the network. Let $x\triangleq(x_1,\ldots,x_M)$ denote a configuration of all these components, where $x_m=1$ represents the operational state and $x_m=0$ represents the failed state of component $m\in[M]$. Each configuration x occurs with a probability $P(x)$. The rare threat $\mathcal{E}$ is represented by a set of profiles x, which is often unknown and of a large cardinality, say $O(2^M)$. Given x, an indicator function $\mathbf{1}_{\mathcal{E}}(x)$ can simulate the network to test the occurrence of $\mathcal{E}$ (i.e., $x\in\mathcal{E}$), but have no functional description of $\mathcal{E}$. Thus, as many as $O(2^M)$ enumerations are needed to evaluate $\mathcal{E}$'s occurrence probability. For more background readings related to this example, readers can refer to static network reliability literatures [3], [4], [5], [6], [7], [8].

A typical method to address the high computational cost issue, as illustrated in Example 1, is Monte Carlo (MC) sampling. It estimates the occurrence of a rare threat $\mathcal{E}$ via generating samples x from $P(x)$ and obtaining simulation results from $\mathbf{1}_{\mathcal{E}}(x)$. However, to obtain accurate estimations, MC needs to simulate plenty of samples x so to capture sufficient occurrences of $\mathcal{E}$ (i.e., $\mathbf{1}_{\mathcal{E}}(x)=1$). To improve the estimation efficiency of MC, importance sampling (IS) takes a customized importance distribution $Q(x)$ to "*boost*" the occurrence of $\mathcal{E}$. One limitation of IS is that it has to simulate each rare threat with its customized importance distribution individually. This leads to an excessively high simulation cost, especially when dealing with *a large number of rare threats* as shown below:

Example 2. Consider N critical flows in Example 1. Let the rare threat $\mathcal{E}_n$ be the failure to provide promised QoS guarantees for flow n, where $n\in[N]$. Each $\mathcal{E}_n$ is associated with a customized importance distribution $Q_n(x)$ and an indicator function $\mathbf{1}_{\mathcal{E}_n}(x)$ that tests whether $x\in\mathcal{E}_n$. Suppose, using IS to estimate each $\mathcal{E}_n$ requires T samples from $Q_n(x)$. To estimate all N events, we need TN samples, which is expensive for a large scale network because even simulating a single sample can take many hours.

A straightforward idea to reduce such simulation cost burden of IS is to design a simulation distribution Q that works efficiently and accurately for every rare threat. Therefore, we consider the *mixture importance sampling (MIS)* with a mixture parameter $w = (w_1, w_2, \dots, w_N)$

$$Q(x;w) = \sum_{n\in[N]} w_n Q_n(x),$$

where $w_n \geq 0, n\in[N]$ and $\sum_{n\in[N]} w_n = 1$. Through this, each sample x drawn from distribution $Q(x;w)$ can be used for "*all*" rare threats $\{\mathcal{E}_n\}_{n=1}^N$. We aim to answer:

- *How to quantify the "simulation cost" for a mixture w?*
- *How to locate the optimal mixture w^*?*

The design of proper simulation cost metrics for w requires the careful consideration of simulation cost resulted from $Q(x;w)$ for each $\mathcal{E}_n$. Such metrics (i.e., cost measures) are functions of $Q(x;w)$. To search for w^* minimizing the simulation cost, one needs to evaluate the search direction by marginalizing x in the metric with a sample space of size 2^M. To address this challenge, we formulate a multi-armed bandit (MAB) online learning (OL) framework to estimate the "*search direction*" and learn w^* from simulation samples x drawn from $\{Q_n(x)\}_{n=1}^N$. One may use the classical stochastic optimization (SO) method to derive w^*, but it has a much more expensive learning cost (i.e., the number of samples x generated to learn w^*) than our framework: in the learning phase, to guarantee a fast convergence to w^*, SO needs sufficient samples from $Q(x;w^{(t)})$ to locate an efficient "*search direction*" in each round t, and $w^{(t)}$ is the estimated mixture; and in the estimation phase, additional samples from $Q(x;w^{(T)})$ are required to provide accurate estimation for $\{\mathcal{E}_n(x)\}_{n=1}^N$. In contrast, our framework only needs a single sample x generated from one of $\{Q_n(x)\}_{n=1}^N$ in each round of learning, and the generated samples can be directly used for providing accurate estimation for $\{\mathcal{E}_n(x)\}_{n=1}^N$. This leads to a "zero learning cost", but also makes it challenging to estimate the "*search direction*" as well as learn w^*. The contributions of our work are:

- We propose *two metrics* to quantify the simulation cost for a mixture strategy and propose *a mixture importance sampling optimization problem (MISO)*to select the optimal mixture. We show the search direction of mixture is costly to evaluate, making it challenging to locate the optima.
- We formulate *a MAB OL framework* which estimates the search direction and learn the optimal mixture from "simulation samples". Instead of using sufficient simulation samples from $Q(x;w^{(t)})$, our framework reduces the simulation cost by generating only a single simulation sample x from one of $\{Q_n(x)\}_{n=1}^N$ in each round of learning. Hence, achieving a zero cost on extra samples.
- We develop *MAB learning algorithms* to efficiently learn w^* under different cost measures, i.e.: (1) SumVar, to minimize the sum of variances with a regret of $(\ln T)^2/T$, and (2) SimCos, to minimize the simulation cost with a regret of $\sqrt{\ln T/T}$, where T is the number of samples. For each algorithm, we provide: (a) convexity and smoothness analysis; (b) algorithm to estimate the search direction of w with zero cost on extra samples, as well as provable concentration; (c) regret analysis and reveal the impact of key factors, e.g., similarity of $\{Q_n(x)\}_{n=1}^N$, on the regret.
- To demonstrate the versatility of our framework, we apply it to conduct rare event simulation over an Internet backbone network and a smart grid network, both of which are supported by real-world data.

2 Problem Formulation

We start by introducing the MIS model with a mixture parameter w. Then, we formulate an optimization framework that minimizes a general cost measure via selecting a proper w. To address the computational challenge in locating the optimal mixture w^*, we develop an OL framework to estimate w^*. Finally, we present two important instances of the OL framework with specific cost measures.

2.1 Mixture Importance Sampling

Assume there are N rare events and each event is induced by a subset of M potential *causes* denoted by $[M]$. We aim to estimate the occurrence probability for each individual event. Let $\Omega \triangleq \{0,1\}^M$. We denote $x = (x_1, \dots, x_M) \in \Omega$ as the *occurrence profile* of all M causes, where x_m indicates whether the cause m occurs (1: yes, 0: no). Let x occur with a probability $P(x)\in [0,1]$, where $\sum_{x\in\Omega} P(x) = 1$. Denote the event $n\in[N]$ as $\mathcal{E}_n\subset\Omega$, of which the occurrence is indicated by a *membership oracle*

$$\mathbf{1}_{\mathcal{E}_n}(x) \triangleq \begin{cases} 1, & \text{if } x \in \mathcal{E}_n, \\ 0, & \text{otherwise}. \end{cases} \quad (1)$$

Notice that $\mathbf{1}_{\mathcal{E}_n}(x)$ can be used with causes indicated by sample x to test whether $\mathcal{E}$ occurs, i.e., $x \in \mathcal{E}$, as long as x is provided. Yet, $\mathbf{1}_{\mathcal{E}_n}$ has no other functional description of $\mathcal{E}_n$. And the occurrence probability is denoted by

$$\mu_n = \mathbb{P}_{x\sim P}[\mathbf{1}_{\mathcal{E}_n}(x) = 1] = \sum_{x\in\mathcal{E}_n} P(x). \quad (2)$$

In many real-life applications, $\mathcal{E}_n$ has a large cardinality, which makes the exact value of μ_n computationally expensive to evaluate. For instance, consider an Internet-scale network with M physical links, where the mth link fails with a probability of p_m. There are N competing flows, of which the undelivery of the nth flow is represented by $\mathcal{E}_n\subset\Omega$. For each $x\in\Omega$, the probability to obverve the occurrence of x in the real-life can be expressed as

$$P(x) = \prod_{m\in[M]} p_m^{x_m}(1-p_m)^{1-x_m}. \quad (3)$$

Due to the high complexity of traffic engineering, $\mathcal{E}_n$ is usually unknown and with a large cardinality, resulting in a computational complexity of $O(2^M)$ to evaluate the exact value of μ_n. Eq. (3) assumes the independence of potential causes, and this assumption is also considered in many previous works [3], [4], [5], [6], [7], [9]. We are also aware of notable works [10], [11] dealing with dependent potential causes. Dependent potential causes greatly complicate the problem and we leave it as our future work.

As it is rare to see the occurrence of $\mathcal{E}_n$, it is costly to estimate μ_n via simulating x with $P(x)$, i.e., the classical MC method. One typical method to address this challenge is the IS method [12], [13]. To improve the efficiency of MC, IS replaces the sampling distribution $P(x)$ with $Q_n(x)$ to increase the occurrence of event $\mathcal{E}_n$, and assigns each sample x a weight to recover the unbiasedness. Specifically, it replaces Eq. (2) by

$$\begin{aligned}\mu &= \mathbb{E}_{x\sim P}[\mathbf{1}_{\mathcal{E}_n}(x)=1]\\ &= \int \mathbf{1}_{\mathcal{E}_n}(x)\tfrac{P(x)}{Q_n(x)}Q_n(x)\mathrm{d}x = \mathbb{E}_{x\sim Q_n}\Big[\mathbf{1}_{\mathcal{E}_n}\tfrac{P(x)}{Q_n(x)}\Big].\end{aligned} \tag{4}$$

Assume each $\mathcal{E}_n$ corresponds to a *customised* pure importance distribution $Q_n(x)$. IS provides an efficient estimation of μ_n if taking $Q_n(x)$ to simulate x, but $Q_n(x)$ may not work for other rare events. The *one-run variance* for estimating μ_n with $Q_n(x)$ to simulate x is

$$\mathbb{V}_{x\sim Q_n}\big[\mathbf{1}_{\mathcal{E}_n}(x)\tfrac{P(x)}{Q_n(x)}\big] \triangleq \mathbb{E}_{x\sim Q_n}\big[\mathbf{1}_{\mathcal{E}_n}(x)\tfrac{P^2(x)}{Q_n^2(x)}\big] - \mu_n^2. \tag{5}$$

Note that the one-run variance is an essential metric to measure the estimation efficiency, and it also determines the simulation cost. Here $\{Q_n(x)\}_{n=1}^N$ can be obtained utilizing IS or Sequential IS methods proposed in [9].

However, given a limited simulation budget and a large N, one usually could not afford to estimate each μ_n individually with the corresponding $Q_n(x)$. What one needs is an efficient sampling distribution working for *multiple* interested events simultaneously. Assume we take a mixture of $\{Q_n(x)\}_{n=1}^N$ as the importance distribution to simulate x. Formally, we have

$$Q(x;w) \triangleq \textstyle\sum_{n\in[N]} w_n Q_n(x), \tag{6}$$

where $w\triangleq(w_1,\dots,w_N), w_n\geq 0$ and $\sum_{n\in[N]}w_n = 1$. For the ease of presentation, denote the set of all possible choices of w as the probability simplex $\Delta\triangleq\{w|w_n\geq 0, \sum_{n=1}^N w_n = 1\}$. Here, we define the "*$\xi$-similarity*" as a metric to quantify how well the occurrences of interested events $\{\mathcal{E}_n\}_{n=1}^N$ can be efficiently estimated together by the following definition:

Definition 1 (. ξ-similarity) *Events $\{\mathcal{E}_n\}_{n=1}^N$ are ξ-similar if their corresponding pure importance distributions $\{Q_n(x)\}_{n=1}^N$ satisfy: for $\xi\in[1,\infty]$, $\forall x\in\boldsymbol{\Omega}, \forall n, n'\in[N], \frac{1}{\xi}\leq\frac{Q_n(x)}{Q_{n'}(x)}\leq\xi$.*

To illustrate, consider $\{Q_n(x)\}_{n=1}^N$ have different (or even disjoint) supports, then $\xi=\infty$. Fig 1 shows more examples with different levels of ξ-similarities.

2.2 General Optimization and Learning Framework

Given the mixture w, we take $Q(x;w)$ to simulate x, and the one-run variance of $\mathcal{E}_n$ is

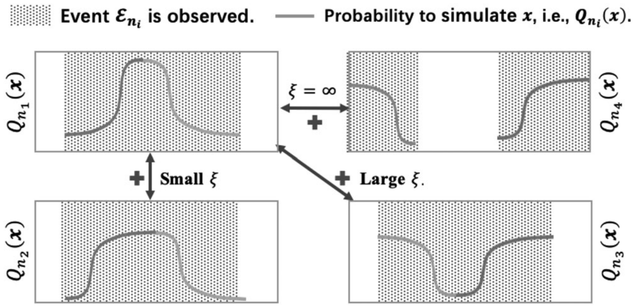


Fig. 1. Examples of different levels of ξ-similarities: an infinite ξ happens when events have different supports and it implies that even the optimal mixture distribution $Q(x;w^*)$ would not work for all events, e.g., $\{\mathcal{E}_{n_1},\mathcal{E}_{n_4}\}$; a large ξ implies a slow convergence to $Q(x;w^*)$, e.g., $\{\mathcal{E}_{n_1},\mathcal{E}_{n_3}\}$; a small ξ implies a fast convergence to $Q(x;w^*)$, e.g., $\{\mathcal{E}_{n_1},\mathcal{E}_{n_2}\}$.

$$\sigma_n^2(w) \triangleq \mathbb{V}_{x\sim Q}\big[\mathbf{1}_{\mathcal{E}_n}(x)\tfrac{P(x)}{Q(x;w)}\big]. \tag{7}$$

One can evaluate the overall simulation efficiency associated with the mixture parameter w by the cost measure $L(\boldsymbol{\sigma}(w))\in\mathbb{R}$ where $\boldsymbol{\sigma}(w)\triangleq(\sigma_1(w)\sigma_2(w),\dots,\sigma_N(w))$ (Refer to Section 2.3 for some examples). We now formulate the mixture importance sampling optimization problem as follows:

Problem 1 (. Mixture Importance Sampling Optimization (MISO) Problem) *Given M causes, associated with a natural occurrence distribution $P(x)$; N interested events, associated with efficient pure importance distributions $\{Q_n(x)\}_{n=1}^N$; and the cost measure $L(\boldsymbol{\sigma}(w))$. Select w to minimize the cost*

$$\min\nolimits_{w\in\Delta} L(\boldsymbol{\sigma}(w)). \tag{8}$$

Problem 1 is essentially a non-linear optimization problem. It is challenging to address it because both the objective $L(\boldsymbol{\sigma}(w))$ and its gradient $\nabla L(\boldsymbol{\sigma}(w))$ are computationally expensive to evaluate: The exact computational complexities are $O(2^M)$ for the large state space of x. To overcome the challenge, we develop a MAB framework to estimate (or online learn) w^* from simulation samples.

Problem 2 (. Mixture Importance Sampling Learning (MIS Learning) Problem) *Given M causes, N interested events and the number of rounds (or data samples) $T\in\mathbb{N}_+$. At round $t=1,\dots,T$:*

- *Select an arm (or event) $I_t\in[N]$ based on an algorithm $\mathcal{A}$ and the sample history $\{(I_s, x^{(s)})\}_{s=1}^{t-1}$;*
- *Draw a simulation sample of profile $x^{(t)}$ from $Q_{I_t}(x)$;*
- *Update the proportions of selecting arms (or events) which denoted by $w^{(t)}=(w_1^{(t)},\dots,w_N^{(t)})$, where $w^{(t)}=\frac{1}{t}\sum_{s\in[t]}e_{I_s}$;*

Objective: Design an MAB algorithm $\mathcal{A}$ to achieve a low and sublinear regret, where the regret is defined as

$$R_T \triangleq L(\boldsymbol{\sigma}(w^{(T)})) - \min\nolimits_{w\in\Delta} L(\boldsymbol{\sigma}(w)). \tag{9}$$

In Problem 2, each arm (or event) indexed by n corresponds to a customized pure distribution $Q_n(x)$, and a general cost function $L(\boldsymbol{\sigma}(w^{(T)}))$ is considered. We next consider two important instances of the MIS-Learning Problem.

2.3 Two Instances of the MIS Learning Problem

Given $Q(x;w)$ to simulate x, let $\ell_n(w)$ measure the simulation cost to achieve the desired estimation accuracy for μ_n, i.e., the confidence interval (CI) is bounded by a threshold δ_n. Also, let $\ell_{max}(w)$ measure the simulation cost to achieve desired estimation accuracies for all $\{\mu_n\}_{n=1}^N$. We have

$$\ell_n(w) \triangleq \frac{\sigma_n^2(w)}{\delta_n^2} \quad \text{and} \quad \ell_{max}(w) \triangleq \max_{n\in[N]}\ell_n(w). \tag{10}$$

Next, we consider various accuracy requirements $\{\delta_n\}_{n=1}^N$, i.e., homogeneous and heterogenous, and study the impact on the cost measure $L(\boldsymbol{\sigma}(w))$. Then, we introduce the corresponding MIS-Learning problems.

MIS-Learning to Minimize the Sum of Variances. We start with the simplest case with homogeneous accuracy requirements (i.e., $\{\delta_n\}_{n=1}^N$ are equal) and consider bounding $\sum_{n\in[N]}\ell_n(w)$ in order to bound $\ell_{max}(w)$. Then

$$\begin{aligned}\min_{w\in\Delta} \textstyle\sum_{n\in[N]}\ell_n(w) &\Longleftrightarrow \min_{w\in\Delta} \textstyle\sum_{n\in[N]}\sigma_n^2(w) \\ &\Longleftrightarrow \min_{w\in\Delta} \textstyle\sum_{n\in[N]}\sigma_n^2(w)+\mu_n^2.\end{aligned} \tag{11}$$

And the total loss (i.e., cost measure) can be defined in terms of the sum of one-run variances as follows:[1]

$$L(\boldsymbol{\sigma}(w)) = \textstyle\sum_{n\in[N]}\sigma_n^2(w)+\mu_n^2 \triangleq L_{\text{SumVar}}(w). \tag{12}$$

We name the MIS-Learning with cost measure in Eq. (12) as minimizing the sum of variances (SumVar) MIS-Learning.

MIS-Learning to Minimize the Simulation Cost. We consider the case where $\{\mathcal{E}_n\}_{n=1}^N$ have heterogenous accuracy requirements. Specifically, we assume each $\mathcal{E}_n$ has a predefined occurrence probability threshold o_n, e.g., $\mathcal{E}_n$ represents the undelivery of a specific flow and we want to *accurately state whether the undelivery probability $\mu_n \le o_n$ or not*. Then the CI width should not exceed $\delta_n = |\mu_n - o_n|$ and

$$\min_{w\in\Delta}\ell_{max}(w) \Longleftrightarrow \min_{w\in\Delta}\max_{n\in[N]} \frac{\sigma_n^2(w)}{(\mu_n - o_n)^2}. \tag{13}$$

The total loss can be defined in terms of the simulation cost to achieve all desired estimation accuracies as

$$L(\boldsymbol{\sigma}(w)) = \max_{n\in[N]} \frac{\sigma_n^2(w)}{(\mu_n - o_n)^2} \triangleq L_{\text{SimCos}}(w). \tag{14}$$

We then name the MIS-Learning with cost measure in Eq. (14) as minimizing the simulation cost (SimCos) MIS-Learning.

3 Learning to Minimize Sum of Variances

We first present our SumVar algorithm design for learning the optimal mixture w^* that minimizes the sum of variances in an online manner. Then we prove a regret upper bound for the SumVar algorithm and reveal the impact of ξ-similarity on the convergence speed to learn w^*.

1. As the analytic expressions of $L_{\text{SumVar}}(w)$ and of its gradient are frequently used in this work, and considering $\sigma_n^2(w) = \mathbb{E}_{x\sim Q}\left[\mathbf{1}_{\mathcal{E}_n}(x)\frac{P^2(x)}{Q^2(x;w)}\right] - \mu_n^2$, we define $L_{\text{SumVar}}(w)$ as illustrated in Eq. (12) for the ease of presentation.

3.1 The Design of SumVar Algorithm

The SumVar algorithm's main idea is that at each round of learning: (1) *First estimate the gradient $\nabla L_{SumVar}(w)$ from historical data samples*; (2) *Then select the arm (or event) based on the estimated gradient*.

Gradient Estimation. We first consider the estimation of gradient $\nabla L_{SumVar}(w^{(t-1)})$ utilizing historical data samples, at each learning round t. We can derive $\nabla L_{SumVar}(w)$ as follows:

$$\begin{aligned}\nabla L_{SumVar}(w) &= \nabla\left\{ \textstyle\sum_{n\in[N]} \mathbb{E}_{x\sim Q(x;w)}\left[\frac{P^2(x)\mathbf{1}_{\mathcal{E}_n}(x)}{Q^2(x;w)}\right]\right\} \\ &= -\textstyle\sum_{n\in[N]}\sum_{x\in\Omega}\left[\frac{P^2(x)\mathbf{1}_{\mathcal{E}_n}(x)}{Q^2(x;w)}\right](Q_1(x),\ldots,Q_N(x)) \\ &= \mathbb{E}_{x\sim Q(x;w)}\left[(-Z_1(x),\ldots,-Z_N(x))\right],\end{aligned} \tag{15}$$

where $Z_n(x) \triangleq \frac{P^2(x)\sum_{i\in[N]}\mathbf{1}_{\mathcal{E}_i}(x)}{Q^3(x;w^{(t-1)})}Q_n(x), \forall n\in[N]$. If historical data samples $\{x^{(s)}\}_{s=1}^{t-1}$ were IID samples of $x\sim Q(x;w^{(t-1)})$, then the gradient $\nabla L_{SumVar}(w^{(t-1)})$ can be estimated by $g^{(t)}$, where

$$g_n^{(t)} = \frac{-1}{t-1}\textstyle\sum_{s\in[t-1]} Z_n(x^{(s)}), \forall n\in[N]. \tag{16}$$

Nevertheless, the challenge is that $\{x^{(s)}\}_{s=1}^{t-1}$ are generated from $x^{(s)}\sim Q_{I_s}(x)$. To address this challenge, the following theorem proves that Eq. (16) is asymptotically accurate in estimating the gradient $\nabla L_{SumVar}(w^{(t-1)})$.

Algorithm 1. SumVar MIS-Learning

Input: N, $w = (\frac{1}{N},\ldots,\frac{1}{N})$, $c_n^{(t)}, \forall n\in[N], t = 1,\ldots,T$
for all $t \le N$ **do**
  Draw $x^{(t)}$ according to $Q_t(x)$, and record history I_t, $Q_n(x^{(t)})$ and $\mathbf{1}_{\mathcal{E}_n}(x^{(t)}), n\in[N]$ for updating $w^{(t)}$ and gradient estimation.
for all $t > N$ **do**
  Estimate the gradient $\nabla L_{SumVar}(w^{(t-1)})$ using $g^{(t)}$ in Eq. (16).
  Compute the lower confidence bound (LCB) $\underline{g}^{(t)}$, where $\underline{g}_n^{(t)} = g_n^{(t)} - c_n^{(t)}$.
  Select $I_t \in \arg\min_{n\in[N]}\underline{g}_n^{(t)}$ and draw $x^{(t)}$ from $Q_{I_t}(x)$.
  Record history I_t, $Q_n(x^{(t)})$ and $\mathbf{1}_{\mathcal{E}_n}(x^{(t)}), n\in[N]$.
  Update $w^{(t)} \leftarrow w^{(t-1)} + \frac{1}{t}(e_{I_t} - w^{(t-1)})$.

Theorem 1. *Consider the MIS-Learning framework, where at round t, $t\in[T]$ take the I_tth distribution $Q_{I_t}(x)$ to generate $x^{(t)}$. Then, $\lim_{t\to\infty}\|g^{(t)} - \nabla L_{SumVar}(w^{(t-1)})\| = 0$.*

Remark. Such asymptotic property owns much to the role of mixture parameter $w^{(t)}$, i.e., the observed proportions of selecting the distribution $Q_{I_t}(x)$ till round t. Hence, after sufficient t rounds of MIS-Learning, all samples $\{x^{(s)}\}_{s=1}^t$ can be approximately considered as simulated by $Q(x;w^{(t)})$.

Arm Selection. Now we outline how the SumVar algorithm in Algorithm 1 selects arm at each learning round. From [14], we notice that finding the minimizer of lower bound confidence $\min_{n\in[N]}\underline{g}_n^{(t)}$ is equivalent to making a step of size $\frac{1}{t+1}$ in the direction of corner of simplex Δ that $\min_{z\in\Delta} z^\top \underline{g}^{(t)}$, which is precisely the Frank-Wolfe algorithm [15]. Hence, we apply the LCB Frank-Wolfe algorithm to select the arm based on the estimated gradient in Eq. (16).

Note that in Algorithm 1, one can select $c_n^{(t)}$ to control the exploration and exploitation tradeoffs.[2]Selecting the $c_n^{(t)}$ is closely related to the regret of Algorithm 1. We thus defer the selection in the next subsection, where we analyze the regret upper bound.

3.2 Main Result on the Regret of SumVar Algorithm

We start by establishing two building blocks for the regret analysis of Algorithm 1: (1) *The strong convexity and smoothness properties of*$L_{\mathrm{SumVar}}(w)$, and (2) *The concentration property of $g^{(t)}$ in estimating $\nabla L_{\mathrm{SumVar}}(w^{(t-1)})$*. Then, we apply these two building blocks to derive the regret upper bound of Algorithm 1.

Strong Convexity and Smoothness of $L_{\mathrm{SumVar}}(w)$. Let us first formally define the strong convexity and smoothness.

Definition 2 (Strong convexity and smoothness). *Let X be a convex set in the vector space and f: $X\to\mathbb{R}$ be a function. Also let I represent the identity matrix. f is called α-strongly convex if and only if $\forall x\in X, \nabla^2 f(x)\succeq\alpha I$, or equivalently*

$$\forall x, y\in X,\ f(y)\geq f(x) + \nabla^{\top} f(x)(y-x) + \tfrac{\alpha}{2}\|y-x\|_2^2. \tag{17}$$

Similarly, f is β-smooth if and only if $\forall x\in X, \nabla^2 f(x)\preceq\beta I$, or equivalently

$$\forall x, y\in X,\ f(y)\leq f(x) + \nabla^{\top} f(x)(y-x) + \tfrac{\beta}{2}\|y-x\|_2^2. \tag{18}$$

The following theorem gives the strong convexity and smoothness of $L_{\mathrm{SumVar}}(w)$.

Theorem 2. *If $\{\mathcal{E}_n\}_{n=1}^N$ has a ξ-similarity, the $L_{\mathrm{SumVar}}(w)$ given by Eq. (12) is α-strongly convex and β-smooth with*

$$\alpha = \tfrac{2}{N\xi^2}\left(\textstyle\sum_{n\in[N]}\mu_n\right)^2 \text{ and } \beta = 2\xi^3\textstyle\sum_{n\in[N]}\mu_n. \tag{19}$$

Remark. Theorem 2 quantifies the impact of ξ-similarities on $L_{\mathrm{SumVar}}(w)$. In particular, the strong convexity of $L_{\mathrm{SumVar}}(w)$ vanishes and $L_{\mathrm{SumVar}}(w)$ becomes nonsmooth when $\xi\to\infty$, i.e., the event occurrences are not similar. This implies that the ξ-similarity is essential for learning the optimal mixture w^* as well. Besides, in case that $\xi=\infty$, one can divide $\{\mathcal{E}_n(x)\}_{n=1}^N$ into multiple sub-groups such that each sub-group has a finite ξ.

Concentration Property of $g^{(t)}$. The preciseness and efficiency of the gradient estimator $g^{(t)}$ directly affect Algorithm 1's performance in minimizing the objective. We aim to characterize how well $g_n^{(t)}$ concentrates around $\nabla L_{\mathrm{SumVar}}(w^{(t)})\big|_n$. Such concentration is characterized by a balance between the confidence probability denoted by $\zeta^{(t)}\in[0,1]$ and the deviation denoted by $\epsilon_n^{(t)}$. One challenge is that in the estimator $g_n^{(t)}$ in Eq. (16), the historical data samples $\{x^{(s)}\}_{s=1}^{t-1}$ are not IID. The following theorem resolves this challenge by quantifying the tradeoff between $\zeta^{(t)}$ and $\epsilon_n^{(t)}$.

2. In Algorithm 1, the derivation of c_n relies on Z_n and its moments, which are costly to compute exactly. In the implementation, we take the empirical estimation of Z_n and its moments. This will not affect our regret upper bound conclusion as the derivation utilizes the upper bounds of Z_n and its moments.

Theorem 3. *Assume $x\sim Q\big(x; w^{(t-1)}\big)$ for both $\mathbb{E}$ and $\mathbb{V}$. Suppose $\zeta^{(t)}$ and $\epsilon^{(t)}$ satisfy*

$$\epsilon_n^{(t)} = \tfrac{1}{3t}\ln\tfrac{1}{\zeta^{(t)}} Z_n^{max} + \sqrt{\tfrac{1}{9t^2}\left(\ln\tfrac{1}{\zeta^{(t)}} Z_n^{max}\right)^2 + \tfrac{2}{t}\ln\tfrac{1}{\zeta^{(t)}}\mathbb{V} Z_n(x)},$$

where $Z_n^{max}\triangleq\max_{x\in\Omega}|Z_n(x)-\mathbb{E}[Z_n(y)]|$. Then, it holds that

$$\mathbb{P}\big[g_n^{(t)} - \nabla L_{\mathrm{SumVar}}(w^{(t)})\big|_n \geq \epsilon_n^{(t)}\big]\leq\zeta^{(t)}, \tag{20}$$

$$\mathbb{P}\big[g_n^{(t)} - \nabla L_{\mathrm{SumVar}}(w^{(t)})\big|_n \leq -\epsilon_n^{(t)}\big]\leq\zeta^{(t)}. \tag{21}$$

Theorem 3 serves as a building block for one to vary $\zeta^{(t)}$ and $\epsilon^{(t)}$, to attain different confidence and variation tradeoffs. This confidence and variation tradeoff is essential to select the parameter $c_n^{(t)}$ of Algorithm 1 and analyze its regret later. We need to point out that, $Z_n^{max}=O(\xi^3)$ and $\mathbb{V} Z_n(x) = O(\xi^5)$, i.e, the CI width of $g^{(t)}$ is proportional to ξ. This reveals the impact of ξ-similarity on the concentration of gradient estimation.

*Regret Upper Bound.*With the above two building blocks, we now select the parameter $c_n^{(t)}$ for Algorithm 1 and show the regret upper bound. We defer the detailed proof to Section 5.

Theorem 4 (Regret upper bound of SumVar algorithm). *Suppose $\{\mathcal{E}_n\}_{n=1}^N$ has a "ξ-similarity". For MIS-Learning with cost measure $L_{\mathrm{SumVar}}(w)$ in Eq. (12), after T steps of the SumVar algorithm with the choice of $c_n^{(t)}=\epsilon_n^{(t)}$ and*

$$\zeta^{(t)} = \begin{cases} T_0^{-2}, & if\ t\leq T_0,\\ t^{-2}, & if\ t>T_0,\end{cases}$$

the following holds: when $\frac{\ln(1/\zeta^{(t)})}{t}\geq\frac{9\sum_{i\in[N]}\mu_i}{4N\xi}$

$$\mathbb{E}_{x\sim Q}[R_T]\leq C_1\tfrac{1}{T} + C_2\tfrac{\ln T}{T}; \tag{22}$$

otherwise

$$\mathbb{E}_{x\sim Q}[R_T]\leq C_3\tfrac{1}{T} + C_4\tfrac{erf\sqrt{\ln T/2}}{T} + C_5\tfrac{\ln T}{T} + C_6\tfrac{(\ln T)^2}{T}, \tag{23}$$

where, $C_1 = O\Big(\frac{N^2(\ln T_0)^2\xi^6}{\alpha\eta^2} + \frac{N^{3/2}\xi^2\sum_{i\in[N]}\mu_i}{T_0} + \frac{N\ln T_0\beta\xi^3}{\alpha\eta^2}\Big)$,
$C_3 = O\Big(\frac{N^{3/2}\xi^2\sum_{i\in[N]}\mu_i}{T_0} + \frac{N(\ln T_0)^2\beta\xi^2}{\alpha\eta^2}\Big)$, $\quad C_2 = C_5 = O(\beta)$,
$C_4 = O\Big(\frac{\sqrt{N\xi^5}\beta}{\alpha\eta^2}\Big), C_6 = O\Big(\frac{N\xi^5}{\alpha\eta^2}\Big)$.

Remark. Theorem 4 shows that the regret upper bound is proportional to the ξ-similarity. It also reveals that a small ξ implies a fast convergence to the optimal mixture.

4 Learning to Minimize Simulation Cost

We first present our SimCos algorithm design for learning the optimal mixture w^* that minimizes the simulation cost in an online manner. Then we prove a regret upper bound for the SimCos algorithm and reveal the impact of ξ-similarity on the convergence speed to learn w^*.

4.1 The Design of SimCos Algorithm

The SimCos algorithm's main idea is that at each round of learning: (1) *First develop a linear approximation framework to locate the search direction*; (2) *Then design an estimator to estimate the search direction from simulation samples*; (3) *Finally, use the estimated search direction to select the arm*.

Search Direction. Different from the SumVar MIS-Learning, where the object is smooth and the gradient implies the search direction of the mixture. In this case, locating the search direction faces the challenge of non-smooth objective $L_{\mathrm{Sim}Cos}(w)$, which takes the pointwise maximum of functions $\ell_n(w)$. Another constraint is that Problem 2 implies a step size of $1/t$ in updating $w^{(t)}$, i.e.,

$$w^{(t+1)} = \frac{tw^{(t)}+z}{t+1}, \tag{24}$$

where $z\in\Delta$. Namely, to determine the search direction, we first need to determine z. Note that $L_{\mathrm{Sim}Cos}(w)$ is a pointwise maximum function and the linearization of a pointwise maximum function behaves similarly to the linearization of a smooth function [16]. Thus, to measure the potential of z in decreasing $L_{\mathrm{Sim}Cos}(w^{(t)})$, we take a *linearization* of $L_{\mathrm{Sim}Cos}(w)$ at $w = w^{(t)}$

$$\begin{aligned} L_{\mathrm{Sim}Cos}(w^{(t)}; z) &= \max_{n\in[N]}\ell_n(w^{(t)}) + \nabla\ell_n(w^{(t)})^\top(w^{(t+1)} - w^{(t)}) \\ &= \max_{n\in[N]}\ell_n(w^{(t)}) + \nabla\ell_n(w^{(t)})^\top\frac{z-w^{(t)}}{t+1}, \end{aligned} \tag{25}$$

and bound its approximation error in the following lemma:

Lemma 1. $\left|L_{\mathrm{Sim}Cos}(w^{(t)}; z) - L_{\mathrm{Sim}Cos}(w^{(t+1)})\right| = O\left(\frac{\xi^3}{(t+1)^2}\right)$.

Lemma 1 states that the approximation error of linear approximation decreases at a rate of $1/t^2$. This implies that the linear approximation is asymptotically accurate in approximating the $L_{\mathrm{Sim}Cos}(w^{(t+1)})$. Hence, given $w^{(t)}$, we consider the *minimizer of* $L_{\mathrm{Sim}Cos}(w^{(t)}; z)$ as the search direction. Furthermore, the minimum of $L_{\mathrm{Sim}Cos}(w^{(t)}; z)$ can be attained by *the standard direction with steepest decrease*, i.e.,

$$\min_{z\in\Delta}L_{\mathrm{Sim}Cos}(w^{(t)}; z) = \min_{z\in\mathcal{U}}L_{\mathrm{Sim}Cos}(w^{(t)}; z), \tag{26}$$

where $\mathcal{U}\triangleq\{e_1,\ldots,e_N\}$ represents the standard basis. This implies that we can reduce the search space from Δ to $\mathcal{U}$, and simplify estimations of the search direction as we proceed to show. We take such steepest decrease direction as the search direction, and denote it by

$$e_{*_t} = \arg\min_{z\in\mathcal{U}}L_{\mathrm{Sim}Cos}(w^{(t)}; z). \tag{27}$$

Search Direction Estimation. We consider the following equivalent form of the search direction

$$e_{*_t} = \arg\min_{z\in\mathcal{U}}L_{\mathrm{Sim}Cos}(w^{(t)}; z) - L_{\mathrm{Sim}Cos}(w^{(t)}). \tag{28}$$

Such form of search direction is useful to estimate the search direction, for the value of $L_{\mathrm{Sim}Cos}(w^{(t)};z) - L_{\mathrm{Sim}Cos}(w^{(t)})$ shrinks in t. As we will show later, this property enables us to derive better concentration results for the search direction estimation. As the search direction is in the set $\mathcal{U}$, we only need to estimate $\{L_{\mathrm{Sim}Cos}(w^{(t)}; e_n)\}_{n=1}^N$ and $L_{\mathrm{Sim}Cos}(w^{(t)})$ so to locate e_{*_t}. Essentially, we need to estimate $\ell_n(w^{(t)})$ and $\nabla\ell_n(w^{(t)})$ from the data samples $\{x^{(s)}\}_{s=1}^{t-1}$. We have similar challenges as in Section 3.1, i.e., data samples are not IID. We then address these challenges with a similar method: We estimate $L_{\mathrm{Sim}Cos}(w^{(t)}; e_n) - L_{\mathrm{Sim}Cos}(w^{(t)})$ as $g_n^{(t)}$ where $g_n^{(t)}$ is derived as $g_n^{(t)} \triangleq \dot{g}_n^{(t)} - \ddot{g}^{(t)}$ and

$$\begin{aligned} \dot{g}_n^{(t)} &= \max_{i\in[N]}\tfrac{t+1}{t}\widehat{A}_i(w^{(t-1)}) - \tfrac{1}{t}\widehat{B}_i(w^{(t-1)}; n) - \frac{(\widehat{\mu}_i^{(t-1)})^2}{(\widehat{\mu}_i^{(t-1)} - o_i)^2}, \\ \ddot{g}^{(t)} &= \max_{i\in[N]}\widehat{A}_i(w^{(t-1)}) - \frac{(\widehat{\mu}_i^{(t-1)})^2}{(\widehat{\mu}_i^{(t-1)} - o_i)^2}, \\ \widehat{B}_i(w^{(t)}; n) &= \frac{1}{(\widehat{\mu}_i^{(t-1)} - o_i)^2}\,\frac{1}{t}\sum_{s\in[t]}\frac{P^2(x^{(s)})\mathbf{1}_{\mathcal{E}_i}(x^{(s)})}{Q^2(x^{(s)}; w^{(t)})}\,\frac{Q_n(x^{(s)})}{Q(x^{(s)}; w^{(t)})}, \\ \widehat{A}_i(w^{(t)}) &= \frac{1}{(\widehat{\mu}_i^{(t-1)} - o_i)^2}\cdot\frac{1}{t}\sum_{s\in[t]}\frac{P^2(x^{(s)})\mathbf{1}_{\mathcal{E}_i}(x^{(s)})}{Q^2(x^{(s)}; w^{(t)})}. \end{aligned} \tag{29}$$

In the following theorem, we prove that the search direction can be estimated asymptotically accurate.

Theorem 5. *Consider the MIS-Learning framework, where at round $t, t\in[T]$ take distribution $Q_{I_t}(x)$ to generate $x^{(t)}$. Then*

$$\lim_{t\to\infty}\|\dot{g}_n^{(t)} - L_{\mathrm{Sim}Cos}(w^{(t)}; e_n)\| = 0, \tag{30}$$

$$\lim_{t\to\infty}\|\ddot{g}^{(t)} - L_{\mathrm{Sim}Cos}(w^{(t)})\| = 0. \tag{31}$$

Remark. Similar as Theorem 1, such asymptotic property owns much to the mixture parameter $w^{(t)}$.

*Arm Selection.*Now we outline how the SumVar algorithm in Algorithm 2 selects arm at each learning round. Based on $g_n^{(t)}, n\in[N]$, we estimate the steepest search direction using the LCB framework and we outline the arm selection in Algorithm 2. Selecting the parameter $c_n^{(t)}$ is closely related to the regret of Algorithm 2.[3]We thus delay the selection in the next section, where we provide the detailed proofs of the regret.

Algorithm 2. SimCos MIS-Learning

Input: N, $w = (\frac{1}{N},\ldots,\frac{1}{N})$
for all $t\leq N$ **do**
  Draw $x^{(t)}$ according to the $Q_t(x)$ and record history I_t, $Q_n(x^{(t)})$ and $\mathbf{1}_{\mathcal{E}_n}(x^{(t)}), n\in[N]$ for updating $w^{(t)}$ and gradient estimation.
for all $t > N$ **do**
  Estimate $\mu_n^{(t-1)}, n\in[N]$ by $\widehat{\mu}_n^{(t-1)} = \frac{1}{t-1}\sum_{s\in[t-1]}\frac{P(x^{(s)})\mathbf{1}_{\mathcal{E}_n}(x^{(s)})}{Q(x^{(s)}; w^{(s)})}$.
  For all arms $n\in[N]$, compute $g_n^{(t)}$, i.e., the estimated linear approximation of decreasing progress achieved by taking different arms at round t according to Eq. (29).
  Compute the LCB $\underline{g}_n^{(t)}$, where $\underline{g}_n^{(t)} = g_n^{(t)} - c_n^{(t)}$.
  Select arm $I_t\in\arg\min_{n\in[N]}\underline{g}_n^{(t)}$.
  Record history I_t, $Q_n(x^{(t)})$ and $\mathbf{1}_{\mathcal{E}_n}(x^{(t)}), n\in[N]$.
  Update $w^{(t)}\leftarrow w^{(t-1)} + \frac{1}{t}(e_{I_t} - w^{(t-1)})$.

3. The derivation of c_n in Algorithm 2 relies on A_i, B_i, and their moments, which is discussed in the next subsection. In the implementation, we take empirical estimations of these values. This will not affect our regret conclusion as its derivation utilizes the upper bounds of A_i, B_i, and their moments.

4.2 Main Result on the Regret of SimCos Algorithm

To first decompose the regret, denote the optimal mixture as w^*, the optimal search direction as e_{*_t}, and the estimated search direction (i.e., the action direction) as e_{I_t}. Then we decompose the regret as follows:

$$L_{\mathrm{Sim}Cos}(w^{(t+1)}) - L_{\mathrm{Sim}Cos}(w^*) \leq L_{\mathrm{Sim}Cos}\left(\frac{tw^{(t)}+e_{I_t}}{t+1}\right) - L_{\mathrm{Sim}Cos}\left(\frac{tw^{(t)}+e_{*_t}}{t+1}\right) \tag{R1}$$

$$+ L_{\mathrm{Sim}Cos}\left(\frac{tw^{(t)}+e_{*_t}}{t+1}\right) - L_{\mathrm{Sim}Cos}\left(\frac{tw^{(t)}+w^*}{t+1}\right) \tag{R2}$$

$$+ L_{\mathrm{Sim}Cos}\left(\tfrac{tw^{(t)}+w^*}{t+1}\right) - L_{\mathrm{Sim}Cos}(w^*). \tag{R3}$$

This decomposition has three parts. Part $R1$ is the *estimation error*, and is essentially governed by the concentration of $g_n^{(t)}$ in estimating $L_{\mathrm{Sim}Cos}(w^{(t)}; e_n) - L_{\mathrm{Sim}Cos}(w^{(t)})$. Part $R2 + R3$ is the *approximation error*, and is essentially governed by the convexity and smoothness of the objective $L_{\mathrm{Sim}Cos}(w)$. Next, similar as the SumVar case, we first establish two building blocks: (1) *The strong convexity and smoothness properties of $L_{\mathrm{Sim}Cos}(w)$ and its components;* (2) *The concentration property of $g^{(t)}$ in estimating $L_{\mathrm{Sim}Cos}(w^{(t)}; z) - L_{\mathrm{Sim}Cos}(w^{(t)})$*. Then we apply these two blocks to bound the regret of Algorithm 2.

Convexity and Smoothness of $L_{\mathrm{Sim}Cos}(w)$ and its Components. As an immediate consequence of Theorem 2, we can derive the strong convexity and smoothness of $\ell_n(w), n\in[N]$, i.e., the components of $L_{\mathrm{Sim}Cos}(w)$:

Corollary 1. *If $\{\mathcal{E}_n\}_{n=1}^N$ has a ξ-similarity, then $\ell_n(w), n\in[N]$ in Eq. (10) is α_n-strongly convex and β_n-smooth, where*

$$\alpha_n = \frac{2\mu_n^2}{\xi^2(\mu_n - o_n)^2} \text{ and } \beta_n = \frac{2\xi^3\mu_n}{(\mu_n - o_n)^2}. \tag{32}$$

Such convexity and smoothness of $\ell_n(w), n\in[N]$ guarantee the convexity of $L_{\mathrm{Sim}Cos}(w)$:

Corollary 2. *If $\{\mathcal{E}_n\}_{n=1}^N$ has a ξ-similarity, then $L_{\mathrm{Sim}Cos}(w)$ in Eq. (14) is α'-strongly convex, where*

$$\alpha' \triangleq \min_{n\in[N]}\alpha_n = \min_{n\in[N]} \frac{2\mu_n^2}{\xi^2(\mu_n - o_n)^2}. \tag{33}$$

Remark. Corollary 1 and 2 quantify the impact of ξ-similarity on the strong convexity and smoothness of $L_{\mathrm{Sim}Cos}(w)$ and its components. Also note that the tight approximation mentioned in Lemma 1 is guaranteed by the strong convexity and smoothness of $\ell_n(w), n\in[N]$.

Concentration Property of $g_n^{(t)}$. The preciseness and efficiency of the linear approximation decreasing progress estimator $g_n^{(t)}$ directly affect Algorithm 2's performance in minimizing the objective. In the following theorem, we characterize how well $g_n^{(t)}$ concentrates around the $L_{\mathrm{Sim}Cos}(w^{(t)}; e_n) - L_{\mathrm{Sim}Cos}(w^{(t)})$.

Theorem 6. *Assume $x\sim Q(x; w^{(t-1)})$ for both $\mathbb{E}$ and $\mathbb{V}$. For any random variable $X(x)$ define $\widetilde{X}(x) \triangleq X(x) - \mathbb{E}X(x)$ and $\varphi(X(x)) \triangleq \frac{2\ln(8/\zeta^{(t)})}{3t}\max\widetilde{X}(x) + \sqrt{\frac{2\ln(8/\zeta^{(t)})}{t}\mathbb{V}\widetilde{X}(x)}$. Suppose $\zeta^{(t)}$ and $\epsilon^{(t)}$ satisfy:*

$$\begin{cases}\zeta^{(t)} = T_0^{-2}, \epsilon_n^{(t)} = \frac{C_1}{t+1}, & if\ t\leq T_0;\\ \zeta^{(t)} = t^{-2},\ \epsilon_n^{(t)} = \max_{k\in[N]}\frac{1}{t+1}\left(a_k^{(t)} + b_{k,n}^{(t)}\right), & if\ t > T_0;\end{cases}$$

where $A_i(x; w^{(t)}) = \frac{1}{(\mu_i - o_i)^2}\frac{P^2(x)\mathbf{1}_{\mathcal{E}_i}(x)}{Q^2(x; w^{(t)})}$, $B_i(x; w^{(t)}; n) = \frac{1}{(\mu_i - o_i)^2}\frac{P^2(x)\mathbf{1}_{\mathcal{E}_i}(x)Q_n(x)}{Q^3(x; w^{(t)})}$, $a_i^{(t)} = \varphi(A_i(x; w^{(t)}))$, $b_{i,n}^{(t)} = \varphi(B_i(x; w^{(t)}; n))$, $C_1 = \max_{k\in[N]}\frac{2\xi^2\mu_k}{(\mu_k - o_k)^2} + \max_{k\in[N]}\frac{2\xi^3}{(\mu_k - o_k)^2}$.

Then, it holds that

$$\mathbb{P}\left[g_n^{(t)} - \left(L_{\mathrm{Sim}Cos}(w^{(t)}; e_n) - L_{\mathrm{Sim}Cos}(w^{(t)})\right)\leq \epsilon_n^{(t)}\right] \leq \zeta^{(t)},$$
$$\mathbb{P}\left[g_n^{(t)} - \left(L_{\mathrm{Sim}Cos}(w^{(t)}; e_n) - L_{\mathrm{Sim}Cos}(w^{(t)})\right)\geq -\epsilon_n^{(t)}\right] \leq \zeta^{(t)}.$$

Remark. We need to point out that $a_i^{(t)} + b_{i,n}^{(t)} = O\left(\sqrt{\xi^3\frac{\ln(8/\zeta^{(t)})}{t}}\right)$ and $C_1 = O(\xi^3)$. Therefore, Theorem 6 reveals the impact of ξ-similarity on the concentration of estimation.

Regret Upper Bound. With the regret decomposition and above two building blocks, we now select the parameter of Algorithm 2 and prove its regret upper bound. We leave the detailed proof in the later discussion.

Theorem 7 (Regret upper bound of SimCos algorithm). *Suppose $\{\mathcal{E}_n\}_{n=1}^N$ has a "ξ-similarity". For MIS-Learning problem with cost measure $L_{\mathrm{Sim}Cos}$ in Eq. (14), after T steps of the SimCos algorithm, with the choice of $c_n^{(t)} = \epsilon_n^{(t)}$ and*

$$\zeta^{(t)} = \begin{cases} T_0^{-2}, & if\ t\leq T_0,\\ t^{-2}, & if\ t > T_0,\end{cases}$$

the following holds:

$$\mathbb{E}[R_T] \leq O(\xi^3)\frac{1}{T} + O(\beta' + \xi^3)\frac{\ln T}{T} + O(\xi^3)\frac{(\ln T)^2}{T} + O(\xi^{5/2})\sqrt{\frac{\ln T}{T}}. \tag{34}$$

Remark. Theorem 7 shows that the regret upper bound is proportional to the ξ-similarity. It also reveals that a small ξ implies a fast convergence to the optimal mixture.

5 Regret Analysis of MIS-Learning Algorithms

In this section, we present a detailed analysis and proof for regret upper bounds of the SumVar and SimCos algorithms. Due to the page limit, we leave proofs of all claims in the appendix, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TMC.2021.3074920.

5.1 Convexity and Smoothness Analysis

We start with proving Theorem 2, which states the convexity and smoothness of $L_{\mathrm{Sum}Var}$. Then, we apply a similar method to analyze the convexity and smoothness of $L_{\mathrm{Sim}Cos}$.

Proof of Theorem 2. For simplicity, we let $x \sim Q$ represent $x \sim Q(x; w)$ and $x \sim P$ represent $x \sim P(x)$ in the proof. Then

$$\sigma_n^2(w) = \mathbb{V}_{x\sim Q}\left[\mathbf{1}_{\mathcal{E}_n}(x)\frac{P(x)}{Q(x;w)}\right] = \mathbb{E}_{x\sim Q}\left[\mathbf{1}_{\mathcal{E}_n}(x)\frac{P^2(x)}{Q^2(x;w)}\right] - \mu_n^2$$
$$= \sum\nolimits_{x\in\Omega}\mathbf{1}_{\mathcal{E}_n}(x)\frac{P^2(x)}{Q(x;w)} - \mu_n^2. \quad (35)$$

Denote $Q(x) = (Q_1(x), \ldots, Q_N(x))$, the gradient of σ_n^2 becomes

$$\nabla\sigma_n^2 = -\mathop{\mathbb{E}}_{x\sim Q}\mathbf{1}_{\mathcal{E}_n}(x)\frac{P^2(x)}{Q^3(x;w)}Q(x) = -\sum_{x\in\Omega}\mathbf{1}_{\mathcal{E}_n}(x)\frac{P^2(x)}{Q^2(x;w)}Q(x). \quad (36)$$

The gradient of $\nabla(\sum_{n\in[N]}\sigma_n^2)$, which is also the Hessian matrix of $L_{SumVar}(w)$ in Eq. (12), can be derived as

$$H(w) = 2\sum\nolimits_{x\in\Omega}\frac{P^2(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q^3(x;w)}Q(x)Q(x)^\top. \quad (37)$$

(1) For the *convexity*, we have the following:

$$Q(x)\,Q(x)^\top\succeq 0 \Rightarrow H(w)\succeq 0. \quad (38)$$

(2) For the *α-strongly convexity*, we can derive

$$H(w)\succeq\alpha I$$

$$\Leftrightarrow \forall z\in\Delta, \quad z^\top H(w)z\geq\alpha z^\top z = \alpha; \quad (39)$$

$$\Leftrightarrow \forall z\in\Delta, \quad \sum\nolimits_{x\in\Omega}\frac{P^2(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q^3(x;w)}Q^2(x;z)\geq\frac{\alpha}{2}; \quad (40)$$

$$\Leftrightarrow \forall z\in\Delta, \quad \mathbb{E}_{x\sim P}\frac{P(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)}\left\{\frac{Q(x;z)}{Q(x;w)}\right\}^2\geq\frac{\alpha}{2}. \quad (41)$$

By the definition of ξ-similarity, we have $\frac{1}{\xi}\leq\frac{Q(x;z)}{Q(x;w)}\leq\xi$. Then

$$\mathop{\mathbb{E}}_{x\sim P}\frac{P(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)}\left\{\frac{Q(x;z)}{Q(x;w)}\right\}^2\geq\frac{1}{\xi^2}\mathop{\mathbb{E}}_{x\sim P}\frac{P(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)}.$$

Using the Cauchy-Schwarz inequality, we have

$$\sum\nolimits_{x\in\Delta}\frac{P^2(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)}\sum\nolimits_{x\in\Delta}\left[Q(x;w)\sum\nolimits_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)\right]$$
$$\geq\left[\sum\nolimits_{x\in\Delta}\left(P(x)\sum\nolimits_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)\right)\right]^2 = \left(\mathbb{E}_{x\sim P}\sum\nolimits_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)\right)^2 \quad (42)$$
$$=\left(\sum\nolimits_{n\in[N]}\mu_n\right)^2.$$

Note that $\sum_{x\in\Delta}\left[Q(x;w)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)\right]\leq N\sum_{x\in\Delta}Q(x;w)=N$. Combining with Eq. (42), we have the following:

$$\mathop{\mathbb{E}}_{x\sim P}\frac{P(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)} = \sum\nolimits_{x\in\Delta}\frac{P^2(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)}\geq\frac{\left(\sum_{n\in[N]}\mu_n\right)^2}{N}. \quad (43)$$

Hence, we can take α as

$$\alpha = \frac{2}{N\xi^2}\left(\sum\nolimits_{n\in[N]}\mu_n\right)^2. \quad (44)$$

(3) About the *β-smoothness*, it suffice to show

$$H(w)\preceq\beta I$$

$$\Leftrightarrow \forall z\in\Delta, \quad z^\top H(w)z\leq\beta z^\top z = \beta; \quad (45)$$

$$\Leftrightarrow \forall z\in\Delta, \quad \sum\nolimits_{x\in\Omega}\frac{P^2(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q^3(x;w)}Q^2(x;z)\leq\frac{\beta}{2}; \quad (46)$$

$$\Leftrightarrow \forall z\in\Delta, \quad \mathbb{E}_{x\sim P}\frac{P(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)}\left\{\frac{Q(x;z)}{Q(x;w)}\right\}^2\leq\frac{\beta}{2}. \quad (47)$$

By the definition of ξ-similarity, we have

$$\mathop{\mathbb{E}}_{x\sim P}\frac{P(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)}\left\{\frac{Q(x;z)}{Q(x;w)}\right\}^2$$
$$\leq\xi^2\mathop{\mathbb{E}}_{x\sim P}\frac{P(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)}t\leq\xi^2\sum\nolimits_{n\in[N]}\mathop{\mathbb{E}}_{x\sim P}\frac{P(x)\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)}. \quad (48)$$

As $Q_n(x)$ is $\mathcal{E}_n$'s "customized" IS distribution, it simulates $\mathcal{E}_n$'s occurrences more often than the natural distribution $P(x)$ does, i.e., $Q_n(x)\mathbf{1}_{\mathcal{E}_n}(x)\geq P(x)\mathbf{1}_{\mathcal{E}_n}(x)$.[4] Hence, if $\mathbf{1}_{\mathcal{E}_n}(x) = 1$, we have

$$Q(x;w)\geq w_nQ_n(x) + (1-w_n)\frac{1}{\xi}Q_n(x)\geq\frac{1}{\xi}P(x). \quad (49)$$

Therefore, $\frac{P(x)}{Q(x;w)}\mathbf{1}_{\mathcal{E}_n}(x)\leq\xi\mathbf{1}_{\mathcal{E}_n}(x)$, and we have

$$\mathbb{E}_{x\sim P}\frac{P(x)\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)}\leq\xi\mathbb{E}_{x\sim P}\mathbf{1}_{\mathcal{E}_n}(x) = \xi\mu_n, \quad (50)$$

$$\mathbb{E}_{x\sim P}\frac{P(x)\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w)}\left\{\frac{Q(x;z)}{Q(x;w)}\right\}^2\leq\xi^3\sum\nolimits_{n\in[N]}\mu_n\leq\frac{\beta}{2}. \quad (51)$$

We can take β as

$$\beta = 2\xi^3\sum\nolimits_{n\in[N]}\mu_n. \quad (52)$$

This completes the proof of Theorem 2. □

Corollary 1 states the convexity and smoothness of $\ell_n(w)$, $n\in[N]$, i.e., components of $L_{SimCos}(w)$. It can be considered as a special case of Theorem 2 with $N = 1$, i.e., $\{\mathcal{E}_n\}_{n=1}^N$

4. $\mathcal{E}_n$'s optimal IS distribution is $Q_n^*(x){=}P(x)\mathbf{1}_{\mathcal{E}_n}(x)/\mu_n$, where μ_n is very small. Compare to $P(x)$, $Q_n^*(x)$ shifts probabilities from unimportant profiles x (i.e., $\mathbf{1}_{\mathcal{E}_n}(x){=}0$) to important ones (i.e., $\mathbf{1}_{\mathcal{E}_n}(x){=}1$). Thus, $Q_n(x) > P(x)$ if $\mathbf{1}_{\mathcal{E}_n}(x){=}1$. This can be easily satisfied if $Q_n(x)$ is customized for $\mathcal{E}_n$ and so well approximates $Q_n^*(x)$. For the design of $Q_n(x)$, please refer to [9].

contains a single event. The proof follows the same way as Theorem 2.

Corollary 2 reveals the convexity and smoothness of L_{SimCos}. It can be proved using the strong convexity of $\ell_n(w)$ in Corollary 1 and the pointwise maximum property of $L_{\mathrm{SimCos}}(w)$.

5.2 Regret Analysis for SumVar Algorithm

We provide the complete analysis for the regret upper bound of SumVar algorithm and state the derivation of Theorem 4.

Let I_t denote the index of arm selected by $\mathcal{A}$ at round t. Recall that e_{I_s} represents the estimated search direction (i.e., the action direction) at round s, and $w^{(t)} = \frac{1}{t}\sum_{s\in[t]} e_{I_s}$ denotes the empirical distribution of arm selections. We can derive the following recurrence:

$$w^{(t+1)} = \frac{tw^{(t)} + e_{I_{t+1}}}{t+1} = w^{(t)} + \frac{e_{I_{t+1}} - w^{(t)}}{t+1}. \tag{53}$$

Let w^* be the optimal mixture

$$w^* = \arg\min_{w\in\Delta} L_{\mathrm{SumVar}}(w), \tag{54}$$

and define $e_{*_{t+1}}$ as the following minimizer:

$$e_{*_{t+1}} = \arg\min_{z\in\Delta} z^\top \nabla L_{\mathrm{SumVar}}(w^{(t)}), \tag{55}$$

which is also the steepest descent direction of $L_{\mathrm{SumVar}}(w^{(t)})$ *with respect to the standard basis*. Note that $e_{*_{t+1}}$ is our desired search direction, and we estimate it with $e_{I_{t+1}}$ based on historical observations. For convenience, denote

$$\varepsilon^{(t+1)} = \nabla L_{\mathrm{SumVar}}(w^{(t)})^\top (e_{I_{t+1}} - e_{*_{t+1}}). \tag{56}$$

The regret analysis of the SumVar algorithm in Theorem 4 can be divided into the following five steps.

Step 1: By the convexity and smoothness of $L_{\mathrm{SumVar}}(w)$, we first partition the regret R_T and show that

$$R_T = \frac{1}{T}\left[\sum_{s\in[T]}\frac{\beta}{s} + \sum_{s\in[T]}\varepsilon^{(s)}\right] \le \beta\frac{\ln T}{T} + \frac{1}{T}\sum_{s\in[T]}\varepsilon^{(s)}. \tag{57}$$

We first claim the following recurrence:

Claim 1. $(t+1)R_{t+1} \le tR_t + \frac{\beta}{(t+1)} + \varepsilon^{(t+1)}$.

Claim 1 implies the following:

$$(t+1)R_{t+1} \le \sum_{s\in[t+1]}\frac{\beta}{s} + \sum_{s\in[t+1]}\varepsilon^{(s)}. \tag{58}$$

Then step 1 is finished by setting $t+1 = T$.

Step 2: To bound R_T, we consider utilizing the concentration property of $g^{(t)}$ to bound $\sum_{s\in[T]}\varepsilon^{(s)}$.

We start by looking at $c_n^{(t)}$, i.e., the confidence bound of estimating $\nabla L_{\mathrm{SumVar}}(w^{(t)})\big|_n$ with $g_n^{(t)}$, which affects the accuracy of estimating e_{*_t} with e_{I_t} when $n = I_t$. The next claim reveals the relationship between $c_{I_t}^{(t)}$ and $\varepsilon^{(t+1)}$:

Claim 2. *Assume* $c_n^{(t)}$ *satisfies*

$$\mathbb{P}\left[g_n^{(t)} - \nabla L_{\mathrm{SumVar}}(w^{(t)})\big|_n \ge c_n^{(t)}\right] \le \zeta^{(t)}, \tag{59}$$

$$\mathbb{P}\left[g_n^{(t)} - \nabla L_{\mathrm{SumVar}}(w^{(t)})\big|_n \le -c_n^{(t)}\right] \le \zeta^{(t)}. \tag{60}$$

Then with a probability at least $1 - 2\zeta^{(t)}$, $\varepsilon^{(t+1)} \le 2c_{I_{t+1}}^{(t)}$.

Next, we derive the expression of $c_n^{(t)}$. Theorem 3 implies that Eqs. (59) and (60) are satisfied if $c_n^{(t)} = \epsilon_n^{(t)}$, where $\epsilon_n^{(t)}$ is defined in Theorem 3.

Proof of Theorem 3. This can be proved using the *Bernstein Inequality* [17].

Recall that $g_n^{(t)} = \frac{-1}{t-1}\sum_{s\in[t-1]} Z_n(x^{(s)})$ and $\nabla L_{\mathrm{SumVar}}(w^{(t)})\big|_n = -\mathbb{E}Z_n(x)$. Then by $P(x)\mathbf{1}_{\mathcal{E}_n}(x) \le Q_n(x)\mathbf{1}_{\mathcal{E}_n}(x)$ and the definition of ξ-similarity, we have:

$$\begin{aligned}\frac{P(x)\mathbf{1}_{\mathcal{E}_n}(x)}{Q(x;w^{(t-1)})} &\le \frac{Q_n(x)\mathbf{1}_{\mathcal{E}_n}(x)}{w_n^{(t-1)}Q_n(x) + \sum_{i\ne n} w_i^{(t-1)}Q_i(x)} \\ &\le \frac{Q_n(x)\mathbf{1}_{\mathcal{E}_n}(x)}{w_n^{(t-1)}Q_n(x) + (1-w_n^{(t-1)})\frac{1}{\xi}Q_n(x)} \le \xi\mathbf{1}_{\mathcal{E}_n}(x).\end{aligned} \tag{61}$$

Therefore, we obtain an upper bound of $Z_n(x)$ as follows:

$$\begin{aligned}Z_n(x) &= \sum_{i\in[N]}\frac{P^2(x)\mathbf{1}_{\mathcal{E}_i}(x)Q_n(x)}{Q^3(x;w^{(t-1)})} \\ &\le \sum_{i\in[N]}\frac{Q_n^2(x)\mathbf{1}_{\mathcal{E}_i}(x)Q_n(x)}{Q^3(x;w^{(t-1)})} \le \xi^3\sum_{i\in[N]}\mathbf{1}_{\mathcal{E}_i}(x) \le N\xi^3.\end{aligned} \tag{62}$$

Similarly, we can also bound $\mathbb{E}Z_n(x)$ and $\mathbb{E}Z_n^2(x)$

$$\begin{aligned}\mathbb{E}Z_n(x) &= \sum_{i\in[N]}\sum_{x\in\Omega}\frac{P^2(x)\mathbf{1}_{\mathcal{E}_i}(x)Q_n(x)}{Q^2(x;w^{(t-1)})} \\ &\le \sum_{i\in[N]}\sum_{x\in\Omega}\frac{P(x)Q_i(x)\mathbf{1}_{\mathcal{E}_i}(x)Q_n(x)}{Q^2(x;w^{(t-1)})} \le \xi^2\sum_{i\in[N]}\mu_i,\end{aligned} \tag{63}$$

$$\begin{aligned}\mathbb{E}Z_n^2(x) &= \sum_{x\in\Omega}\frac{P^4(x)\left(\sum_{i\in[N]}\mathbf{1}_{\mathcal{E}_i}(x)\right)^2 Q_n^2(x)}{Q^5(x;w^{(t-1)})} \\ &\le \sum_{i,j\in[N]}\sum_{x\in\Omega}\frac{P(x)Q_i^2(x)Q_j(x)\mathbf{1}_{\mathcal{E}_i}(x)\mathbf{1}_{\mathcal{E}_j}(x)Q_n^2(x)}{Q^5(x;w^{(t-1)})} \\ &\le \xi^5\sum_{i,j\in[N]}\sum_{x\in\Omega}\mathbf{1}_{\mathcal{E}_i}(x)\mathbf{1}_{\mathcal{E}_j}(x)P(x) \le N\xi^5\sum_{i\in[N]}\mu_i.\end{aligned} \tag{64}$$

By centering $Z_n(x)$, we have

$$\mathbb{V}[Z_n(x) - \mathbb{E}Z_n(y)] \le N\xi^5\sum_{i\in[N]}\mu_i, \tag{65}$$

$$|Z_n(x) - \mathbb{E}Z_n(y)| \le N\xi^3. \tag{66}$$

Hence, $Z_n^{\max} \le N\xi^3$. For presentation convenience, denote

$$\phi = -\frac{t(\epsilon_n^{(t)})^2}{2Z_n^{\max}\epsilon_n^{(t)}/3 + 2\mathbb{V}Z_n}.$$

Then by the Bernstein inequality, we

$$\mathbb{P}\left[\frac{1}{t}\sum_{s\in[t]}Z_n(x^{(s)}) - \mathbb{E}Z_n(y)\geq\epsilon_n^{(t)}\right]\leq e^{\phi}$$
$$\mathbb{P}\left[\frac{1}{t}\sum_{s\in[t]}Z_n(x^{(s)}) - \mathbb{E}Z_n(y)\leq-\epsilon_n^{(t)}\right]\leq e^{\phi}.$$

We derive ϵ_n and complete the proof by solving $\phi = \ln\zeta^{(t)}$. □

Finally, we bound $c_n^{(t)}$ by the following claim:

Claim 3. *With the choice of $c_n^{(t)} = \epsilon_n^{(t)}$, we have*

$$c_n^{(t)}\leq\begin{cases}\frac{4}{3}N\xi^3\frac{\ln(1/\zeta^{(t)})}{t}, & if\ \frac{\ln(1/\zeta^{(t)})}{t}\geq\frac{9\sum_{m\in[N]}\mu_m}{4N\xi};\\ 2\sqrt{N\xi^5\sum_{m\in[N]}\mu_m\frac{\ln(1/\zeta^{(t)})}{t}}, & otherwise\end{cases}.$$

where $\epsilon_n^{(t)}$ is defined in Theorem 3.

Remark. Combine Claim 2, 3 and Theorem 3, we see that with the choice of $c_n^{(t)}$ in Theorem 3, the regret R_T converges at a rate of $O\left(\frac{1}{T}\sum_t c_{I_t}^{(t)}\right)$ and the bound of $c_{I_t}^{(t)}$ is given by Claim 3.

Step 3: Next, we show that R_T can converge at a faster rate of $O\left(\frac{1}{T}\sum_t \left(c_{I_t}^{(t)}\right)^2\right)$ instead of $O\left(\frac{1}{T}\sum_t c_{I_t}^{(t)}\right)$.

Denote η as the distance from w^* to $\partial\Delta$, i.e., the boundary of Δ. We change the recurrence in Claim 1 as follows:

Claim 4. *Denote $\psi(x) = x^2-\sqrt{2\alpha\eta^2}x$. Then*

$$(t+1)R_{t+1}\leq tR_t+\frac{\left(\varepsilon^{(t+1)}\right)^2}{2\alpha\eta^2}+\frac{\beta}{t+1}+\left[\psi\left(\sqrt{R_t}\right)-\psi\left(\frac{\varepsilon^{(t+1)}}{\sqrt{2\alpha\eta^2}}\right)\right].$$

Let $c^{(t)}\triangleq\max_{n\in[N]}c_n^{(t)}$. Claim 3 can be applied to derive an upper bound for $c_n^{(t)}$ and $c^{(t)}$. We next utilize $c^{(t)}$ and Claim 4 to bound R_T.

Claim 5. *Assume we select $\zeta^{(t)}$ properly such that*

$$\frac{1}{t}\sum_{s\in[t]}(c^{(t)})^2\geq(c^{(t+1)})^2.$$

Then with a probability at least $1-N\sum_t\zeta^{(t)}$ it holds that:

$$TR_T\leq\frac{\alpha\eta^2}{2}+\frac{\pi^2\beta^2}{3\alpha\eta^2}+\beta\ln T+\frac{8\beta}{\alpha\eta^2}\sum_{t\in[T]}\frac{c^{(t)}}{t}+\frac{8}{\alpha\eta^2}\sum_{t\in[T]}(c^{(t)})^2.$$

Step 4: Now we discuss how to select $\zeta^{(t)}$ to guarantee that $\frac{1}{t}\sum_{s\in[t]}(c^{(t)})^2\geq(c^{(t+1)})^2$, and bound $\sum_t(c^{(t)})^2$ and $\sum_t\frac{c^{(t)}}{t}$.

From the previous discussion, Claim 3 gives the upper bound of $c^{(t)}$. Observe that $\frac{1}{t}\sum_{s\in[t]}(c^{(t)})^2\geq(c^{(t+1)})^2$ is achieved if $c^{(t)}$ decreases in t. And we need the lower bound probability $1-N\sum_t\zeta^{(t)}$ to be large enough. We select $\zeta^{(t)}$ and bound $\sum_t(c^{(t)})^2$ and $\sum_t\frac{c^{(t)}}{t}$ by the following:

Claim 6. *With the choice of $\zeta^{(t)} = \begin{cases}T_0^{-2}, & if\ t\leq T_0;\\ t^{-2}, & if\ t>T_0.\end{cases}$*
If $c^{(t)} = \frac{4}{3}N\xi^3\frac{\ln(1/\zeta^{(t)})}{t}$, we have

$$\begin{cases}\sum_{t\in[T]}(c^{(t)})^2\leq\frac{64N^2\xi^6}{9}\left[\frac{\pi^2(\ln T_0)^2}{6}+2\right],\\ \sum_{t\in[T]}\frac{c^{(t)}}{t}\leq\frac{8N\xi^3}{3}\left[\frac{\pi^2\ln T_0}{6}+1\right].\end{cases}$$

If $c^{(t)} = 2\sqrt{N\xi^5\sum_m\mu_m}\sqrt{\frac{\ln(1/\zeta^{(t)})}{t}}$, we have

$$\begin{cases}\sum_{t\in[T]}(c^{(t)})^2\leq4N\xi^5\sum_m\mu_m[(\ln T_0)^2+(\ln T)^2],\\ \sum_{t\in[T]}\frac{c^{(t)}}{t}\leq\sqrt{8N\xi^5\sum_m\mu_m}\{(2+\sqrt{2})\sqrt{\ln T_0}\\ \qquad+\sqrt{2\pi}erf(\sqrt{\ln T/2})\}.\end{cases}$$

Step 5: Now we show the formal regret upper bound of the SumVar algorithm proposed in Theorem 4.

Proof of Theorem 4. Note that

$$\|\nabla L_{SumVar}[w]\|_\infty = \max_{n\in[N],x\in\Delta}|\mathbb{E}Z_n(x)|\leq\xi^2\sum_{i\in[N]}\mu_i. \tag{67}$$

With a probability at most $N\sum_t\zeta^{(t)}\leq\frac{2N}{T_0}$, we have

$$\begin{aligned}&R_T = L_{SumVar}(w) - L_{SumVar}(w^*)\\ &\leq\nabla L_{SumVar}(w)^\top(w-w^*)\leq\|\nabla L_{SumVar}(w)\|_2\|w-w^*\|_2\\ &\leq\sqrt{N}\|\nabla L_{SumVar}(w)\|_\infty\leq\sqrt{N}\xi^2\sum_{i\in[N]}\mu_i.\end{aligned} \tag{68}$$

Also, with a probability at least $1-N\sum_t\zeta^{(t)}$, we have the bound of R_T in Claim 5. By plugging bounds of $\sum_t(c^{(t)})^2$ and $\sum_t\frac{c^{(t)}}{t}$ into Claim 5, we complete the proof. □

5.3 Regret Analysis for SimCos Algorithm

We provide the complete analysis for the regret upper bound of the SimCos algorithm, which learns the optimal mixture w^*

$$w^* = \arg\min_{w\in\Delta}L_{SimCos}(w) = \arg\min_{w\in\Delta}\arg\max_{i\in[N]}\ell_i(w). \tag{69}$$

When updating $w^{(t+1)}$ by $w^{(t+1)} = \frac{tw^{(t)}+z}{t+1}$, the SimCos algorithm first locates the search direction as $z = e_{*_t}$, then estimates e_{*_t} by e_{I_t}. Next, we will show *the reasonability in locating the search direction e_{*_t}* and *the regret bound of e_{I_t} in learning w^**.

(1) Reasonability in Locating the Search Direction.

To locate the search direction z to update $w^{(t+1)}$, we consider $e_{*_t}=\arg\min_{z\in\Delta}L_{SimCos}(w^{(t)};z)$. The reason is that Lemma 1 implies that z's potential in minimizing $L_{SimCos}(w^{(t)};z)$ approximately measures z's potential in decreasing $L_{SimCos}(w^{(t)})$, and the approximation error decrease at a rate of $\frac{1}{t^2}$.

We first give the proof of Lemma 1:

Proof of Lemma 1. Specifically, we can show that

$$L_{SimCos}(w^{(t)};z) - L_{SimCos}(w^{(t+1)})\in\left[-\frac{\beta'}{2},-\frac{\alpha'}{2}\right]\cdot\left\|\frac{z-w^{(t)}}{t+1}\right\|_2^2,$$

where $\alpha' = \min_{n\in[N]}\alpha_n$, $\beta' = \max_{n\in[N]}\beta_n$. As $L_{SimCos}(w)$ is the pointwise maxima of α_n-strong convex and β_n smooth components $\ell_n(w), n\in[N]$, its linearization $L_{SimCos}(w^{(t)};z)$

has the above properties. More details can be found in [16]. □

Let $\mathcal{U} \triangleq \{e_1, \dots, e_N\}$ represent the standard basis. We claim that the search space of e_{*_t} can be reduced from Δ to $\mathcal{U}$, i.e.,:

Claim 7. $\min_{z\in\Delta} L_{SimCos}(w^{(t)};z) = \min_{z\in\mathcal{U}} L_{SimCos}(w^{(t)};z)$.

We reorganize e_{*_t} and thus get the desired search direction in Eq. (28), i.e., $e_{*_t} = \arg\min_{z\in\mathcal{U}} L_{SimCos}(w^{(t)};z) - L_{SimCos}(w^{(t)})$.

(2) Regret Bound in Learning the Optimal Mixture.

To estimate the desired search direction e_{*_t}, the SimCos algorithm estimates $L_{SimCos}(w^{(t)}; e_n) - L_{SimCos}(w^{(t)})$ by $g_n^{(t)}$ and so estimates e_{*_t} by e_{I_t}, $I_t = \arg\min_{n\in[N]} g_n^{(t)}$. The regret analysis of the SimCos algorithm can be divided into six steps.

Step 1: We first decompose the regret as

$$L_{SimCos}(w^{(t+1)}) - L_{SimCos}(w^*) \le R1 + R2 + R3, \tag{70}$$

where part $R1$, $R2$ and $R3$ are given in Section 4.2.

$R1$ is the *estimation error* and $R2 + R3$ is the *approximation error*. We will bound each part of the regret in later discussion.

Step 2: Derive an upper bound for $R1$

$$R1 \le 2c_{I_t}^{(t)} + \frac{\beta'}{2}\left\|\frac{e_{I_t} - w^{(t)}}{t+1}\right\|_2^2 - \frac{\alpha'}{2}\left\|\frac{e_{*_t} - w^{(t)}}{t+1}\right\|_2^2. \tag{71}$$

By the definition of $L_{SimCos}(w^{(t)};z)$ and Lemma1, we have

$$\begin{aligned} R1 \le\ & L_{SimCos}(w^{(t)}; e_{I_t}) - L_{SimCos}(w^{(t)}; e_{*_t}) \\ & + \frac{\beta'}{2}\left\|\frac{e_{I_t} - w^{(t)}}{t+1}\right\|_2^2 - \frac{\alpha'}{2}\left\|\frac{e_{*_t} - w^{(t)}}{t+1}\right\|_2^2. \end{aligned} \tag{72}$$

Hence, to bound $R1$, we focus on bounding $L_{SimCos}(w^{(t)}; e_{I_t}) - L_{SimCos}(w^{(t)}; e_{*_t})$. We first look at $c_n^{(t)}$, i.e., the confidence bound for estimating $L_{SimCos}(w^{(t)}; e_n) - L_{SimCos}(w^{(t)})$ by $g_n^{(t)}$, which affects the accuracy of estimating e_{*_t} by e_{I_t} when $n = I_t$. The relationship between $L_{SimCos}(w^{(t)};e_{I_t}) - L_{SimCos}(w^{(t)};e_{*_t})$ and $c_{I_t}^{(t)}$ can be revealed by the next claim:

Claim 8. *Assume $c_n^{(t)}$ satisfies*

$$\mathbb{P}\left[g_n^{(t)} - \left(L_{SimCos}(w^{(t)};e_{I_t}) - L_{SimCos}(w^{(t)};e_{*_t})\right) \ge c_n^{(t)}\right] \le \zeta^{(t)},$$
$$\mathbb{P}\left[g_n^{(t)} - \left(L_{SimCos}(w^{(t)};e_{I_t}) - L_{SimCos}(w^{(t)};e_{*_t})\right) \le -c_n^{(t)}\right] \le \zeta^{(t)}.$$

Then with a probability at least $1 - 2\zeta^{(t)}$

$$L_{SimCos}(w^{(t)};e_{I_t}) - L_{SimCos}(w^{(t)};e_{*_t}) \le 2c_{I_{t+1}}^{(t)}.$$

Combining Eq.(72) and Claim 8, we finish bounding $R1$.

Step 3: Derive an upper bound for $R2$

$$R2 \le -\frac{\alpha'}{2}\left\|\frac{w^* - w^{(t)}}{t+1}\right\|_2^2 + \frac{\beta'}{2}\left\|\frac{e_{*_t} - w^{(t)}}{t+1}\right\|_2^2. \tag{73}$$

By the optimality of e_{*_t}, we derive that $L_{SimCos}(w^{(t)}; e_{*_t}) \le L_{SimCos}(w^{(t)};w^*)$. Combining with Lemma 1, we have

$$\begin{aligned} L_{SimCos}\left(\frac{tw^{(t)} + e_{*_t}}{t+1}\right) &\le L_{SimCos}(w^{(t)};e_{*_t}) + \frac{\beta'}{2}\left\|\frac{e_{*_t} - w^{(t)}}{t+1}\right\|_2^2 \\ &\le L_{SimCos}(w^{(t)};w^*) + \frac{\beta'}{2}\left\|\frac{e_{*_t} - w^{(t)}}{t+1}\right\|_2^2. \end{aligned}$$
$$L_{SimCos}(w^{(t)};w^*) \le L_{SimCos}\left(\frac{tw^{(t)} + w^*}{t+1}\right) - \frac{\alpha'}{2}\left\|\frac{w^* - w^{(t)}}{t+1}\right\|_2^2.$$

Hence, we finish bounding $R2$

Step 4: Derive an upper bound for $R3$

$$R3 \le \frac{t}{t+1}\left[L_{SimCos}(w^{(t)}) - L_{SimCos}(w^*)\right] + \frac{\beta' - \alpha'}{2}\left\|\frac{w^* - w^{(t)}}{t+1}\right\|_2^2. \tag{74}$$

Let $i = \arg\max_{n\in[N]} \ell_n\left(\frac{tw^{(t)} + w^*}{t+1}\right)$. By the α_i-strongly convexity and β_i-smoothness of $\ell_i(w)$

$$\begin{aligned} L_{SimCos}\left(\frac{tw^{(t)} + w^*}{t+1}\right) &= \ell_i\left(\frac{tw^{(t)} + w^*}{t+1}\right) \\ &\le \ell_i(w^{(t)}) + \nabla\ell_i(w^{(t)})^\top \frac{w^* - w^{(t)}}{t+1} + \frac{\beta_i}{2}\left\|\frac{w^* - w^{(t)}}{t+1}\right\|_2^2 \\ &\le \ell_i(w^{(t)}) + \frac{\ell_i(w^*) - \ell_i(w^{(t)})}{t+1} + \frac{\beta_i - \alpha_i}{2}\left\|\frac{w^* - w^{(t)}}{t+1}\right\|_2^2 \\ &\le \frac{t}{t+1} L_{SimCos}(w^{(t)}) + \frac{1}{t} L_{SimCos}(w^*) + \frac{\beta' - \alpha'}{2}\left\|\frac{w^* - w^{(t)}}{t+1}\right\|_2^2. \end{aligned}$$

Hence, we finish bounding $R3$.

Step 5: Combining the upper bound of each part, we have

$$R_T \le \frac{2}{T}\sum_{t\in[T-1]} c_{I_t}^{(t)} + 3(\beta' - \alpha')\frac{\ln(T/2)}{T}. \tag{75}$$

Now, combining Eqs. (71), (73) and (74), we have

$$\begin{aligned} R_{t+1} \le\ & \frac{t}{t+1}R_t + 2c_{I_t}^{(t)} + \frac{\beta'}{2}\left\|\frac{e_{I_t} - w^{(t)}}{t+1}\right\|_2^2 + \frac{\beta' - \alpha'}{2}\left\|\frac{e_{*_t} - w^{(t)}}{t+1}\right\|_2^2 \\ & + \frac{2\beta' - \alpha'}{2}\left\|\frac{w^* - w^{(t)}}{t+1}\right\|_2^2 \le \frac{t}{t+1}R_t + 2c_{I_t}^{(t)} + \frac{3(\beta' - \alpha')}{(t+1)^2}. \end{aligned} \tag{76}$$

$$\begin{aligned} TR_T &\le 2\sum_{t\in[T-1]}\left[c_{I_t}^{(t)} + \frac{3(\beta' - \alpha')}{t+1}\right] \\ &\le 2\sum_{t\in[T-1]} c_{I_t}^{(t)} + 3(\beta' - \alpha')\ln\frac{T}{2}. \end{aligned} \tag{77}$$

Hence, we complete step 5.

Step 6: We consider bounding $c_n^{(t)}$, which measures the accuracy in estimating $L_{SimCos}(w^{(t)};e_n) - L_{SimCos}(w^{(t)})$ by $g_n^{(t)}$.

In this step, assume $x \sim Q(x;w^{(t)})$ for all $\mathbb{E}$ and $\mathbb{V}$ if unspecified. To finish this step, we introduce the following components:

$$A_i(x;w^{(t)}) \triangleq \frac{1}{(\mu_i - o_i)^2}\frac{P^2(x)\mathbf{1}_{\mathcal{E}_i}(x)}{Q^2(x;w^{(t)})},\ A_i(w^{(t)}) \triangleq \mathbb{E}A_i(x;w^{(t)});$$
$$B_i(x;w^{(t)};n) \triangleq A_i(x;w^{(t)})\frac{Q_n(x)}{Q(x;w^{(t)})},\ B_i(w^{(t)};n) \triangleq \mathbb{E}B_i(x;w^{(t)};n).$$

By definitions of $\ell_i(w^{(t)})$ and $\nabla\ell_i(w^{(t)})^\top(\frac{e_n - w^{(t)}}{t+1})$, we have

$$\ell_i(w^{(t)}) = A_i(w^{(t)}) - \frac{\mu_i^2}{(\mu_i - o_i)^2}, \tag{78}$$

$$\nabla\ell_i(w^{(t)})^\top\left(\frac{e_n - w^{(t)}}{t+1}\right) = \frac{A_i(w^{(t)}) - B_i(w^{(t)};n)}{t+1}. \tag{79}$$

Meanwhile, we have the following unbiased estimators:

$$\widehat{A}_i(w^{(t)}) = \frac{1}{(\mu_i - o_i)^2}\frac{1}{t}\sum\nolimits_{s\in[t]} A_i(x^{(s)};w^{(t)}), \tag{80}$$

$$\widehat{B}_i(w^{(t)};n) = \frac{1}{(\mu_i - o_i)^2}\frac{1}{t}\sum\nolimits_{s\in[t]} B_i(x^{(s)};w^{(t)};n), \tag{81}$$

$$\widehat{\ell}_i(w^{(t)}) = \widehat{A}_i(w^{(t)}) - \frac{\mu_i^2}{(\mu_i - o_i)^2}, \tag{82}$$

$$\widehat{\nabla}\ell_i(w^{(t)})^\top\left(\frac{e_n - w^{(t)}}{t+1}\right) = \frac{\widehat{A}_i(w^{(t)}) - \widehat{B}_i(w^{(t)};n)}{t+1}. \tag{83}$$

We first look at concentrations of $\widehat{A}_i(w^{(t)})$ and $\widehat{B}_i(w^{(t)};n)$, i.e., the key components of $g_n^{(t)}$ in Eq. (29), in the next claims:

Claim 9. *With the choice of* $a_i^{(t)} = \varphi(A_i(x;w^{(t)}))$, *it holds that*

$$\mathbb{P}\left[A_i(w^{(t)}) - \widehat{A}_i(w^{(t)}) \geq a_i^{(t)}\right] \leq \frac{\zeta^{(t)}}{8},$$
$$\mathbb{P}\left[A_i(w^{(t)}) - \widehat{A}_i(w^{(t)}) \leq -a_i^{(t)}\right] \leq \frac{\zeta^{(t)}}{8},$$
$$a_i^{(t)} \leq \frac{2\xi^2}{3(\mu_i - o_i)^2}\frac{\ln(8/\zeta^{(t)})}{t} + \frac{\sqrt{2\xi^3\mu_i}}{(\mu_i - o_i)^2}\sqrt{\frac{\ln(8/\zeta^{(t)})}{t}}.$$

Claim 10. *With the choice of* $b_{i,n}^{(t)} = \varphi\big(B_i(x;w^{(t)};n)\big)$, *then*

$$\mathbb{P}\left[B_i(w^{(t)};n) - \widehat{B}_i(w^{(t)};n) \geq b_{i,n}^{(t)}\right] \leq \frac{\zeta^{(t)}}{8},$$
$$\mathbb{P}\left[B_i(w^{(t)};n) - \widehat{B}_i(w^{(t)};n) \leq -b_{i,n}^{(t)}\right] \leq \frac{\zeta^{(t)}}{8},$$
$$b_{i,n}^{(t)} \leq \frac{2\xi^3}{3(\mu_i - o_i)^2}\frac{\ln(8/\zeta^{(t)})}{t} + \frac{\sqrt{2\xi^5\mu_i}}{(\mu_i - o_i)^2}\sqrt{\frac{\ln(8/\zeta^{(t)})}{t}}.$$

Let the active sets of $L_{SimCos}(w^{(t)};e_n)$ and $L_{SimCos}(w^{(t)})$ be

$$I(w^{(t)};n) = \left\{k \middle| \ell_k(w^{(t)}) + \nabla\ell_k(w^{(t)})^\top\frac{e_n - w^{(t)}}{t+1} = L_{SimCos}(w^{(t)};e_n), k\in[N]\right\},$$
$$I(w^{(t)}) = \left\{k \middle| \ell_k(w^{(t)}) = L_{SimCos}(w^{(t)}), k\in[N]\right\}.$$

Also, define the complement set by

$$I^c\big(w^{(t)};n\big) = [N]\backslash I\big(w^{(t)};n\big) \text{ and } I^c\big(w^{(t)}\big) = [N]\backslash I\big(w^{(t)}\big).$$

To complete analyzing concentration properties of $c_n^{(t)}$ in Theorem 6, we first propose the following statement:

Claim 11. *Let* $i = \arg\max_k \widehat{\ell}_k(w^{(t)}) + \widehat{\nabla}\ell_k(w^{(t)})^\top(\frac{e_n - w^{(t)}}{t+1})$ *and* $j = \arg\max_k \widehat{\ell}_k(w^{(t)})$. *When t is large enough such that:*

$$t \geq \max_{i\neq j}\frac{(\xi - \frac{1}{\xi})\left[A_j(w^{(t)}) + a_j^{(t)}\right] + (1 - \frac{1}{\xi})\left[\frac{\mu_i^2}{(\mu_i - o_i)^2} - \frac{\mu_j^2}{(\mu_j - o_j)^2}\right]}{\left[A_i(w^{(t)}) - a_i^{(t)} - \frac{\mu_i^2}{(\mu_i - o_i)^2}\right] - bigg[A_j(w^{(t)}) + a_j^{(t)} - tfrac\mu_j^2(\mu_j - o_j)^2]}$$
$$- 2 + \frac{1}{\xi}, \tag{84}$$

we have $i = j$ *with a probability at least* $1 - \frac{\zeta^{(t)}}{4}$.

Next, we give the proof of Theorem 6:

Proof of Theorem 6. Let $i' \in I(w^{(t)};n)$, $j' \in I(w^{(t)})$. Also denote $i = \arg\max_k \widehat{\ell}_k(w^{(t)}) + \widehat{\nabla}\ell_k(w^{(t)})^\top(\frac{e_n - w^{(t)}}{t+1})$, $j = \arg\max_k \widehat{\ell}_k(w^{(t)})$.

Case 1: $\epsilon_n^{(t)} = \frac{C_1}{t+1}$. By the optimality of i' and i', we have:

$$\begin{aligned} &L_{SimCos}(w^{(t)};e_n) - L_{SimCos}(w^{(t)}) \\ &= \max_{i'}\ell_{i'}(w^{(t)}) + \nabla\ell_{i'}(w^{(t)})^\top\left(\frac{e_n - w^{(t)}}{t+1}\right) - \max_{j'}\ell_{j'}(w^{(t)}) \\ &\in \left[\frac{A_{j'}(w^{(t)}) - B_{j'}(w^{(t)};n)}{t+1}, \frac{A_{i'}(w^{(t)}) - B_{i'}(w^{(t)};n)}{t+1}\right]; \end{aligned} \tag{85}$$

$$\begin{aligned} &\widehat{L}_{SimCos}(w^{(t)};e_n) - \widehat{L}_{SimCos}(w^{(t)}) \\ &= \max_i\widehat{\ell}_i(w^{(t)}) + \widehat{\nabla}\ell_i(w^{(t)})^\top\left(\frac{e_n - w^{(t)}}{t+1}\right) - \max_j\widehat{\ell}_j(w^{(t)}) \\ &\in \left[\frac{\widehat{A}_j(w^{(t)}) - \hat{B}_j(w^{(t)};n)}{t+1}, \frac{\widehat{A}_i(w^{(t)}) - \hat{B}_i(w^{(t)};n)}{t+1}\right]. \end{aligned} \tag{86}$$

For $\forall k$, we have

$$\begin{aligned}&\left|A_k(w^{(t)}) - B_k(w^{(t)};n)\right| = \left|\nabla\ell_k(w^{(t)})^\top(e_n - w^{(t)})\right|\\ &\leq\frac{1}{(\mu_k - o_k)^2}\xi^2\mu_k\left\|e_n - w^{(t)}\right\|_1\leq\frac{2\xi^2\mu_k}{(\mu_k - o_k)^2},\end{aligned} \tag{87}$$

$$\begin{aligned}&\left|\widehat{A}_k(w^{(t)}) - \widehat{B}_k(w^{(t)};n)\right| = \left|\hat{\nabla}\ell_k(w^{(t)})^\top(e_n - w^{(t)})\right|\\ &\leq\frac{1}{(\mu_k - o_k)^2}\xi^3\left\|e_n - w^{(t)}\right\|_1\leq\frac{2\xi^3}{(\mu_k - o_k)^2}.\end{aligned} \tag{88}$$

Combining Eqs. (85), (86), (87) and (88), we have

$$\begin{aligned}&\left|\left\{L_{SimCos}(w^{(t)};e_n) - L_{SimCos}(w^{(t)})\right\} - g_n^{(t)}\right|\\ &\leq\frac{1}{t+1}\left[\max_{k\in[N]}\frac{2\xi^2\mu_k}{(\mu_k - o_k)^2} + \max_{k\in[N]}\frac{2\xi^3}{(\mu_k - o_k)^2}\right].\end{aligned} \tag{89}$$

Hence, $\epsilon_n^{(t)}\leq\frac{C_1}{t+1}$ always holds. The $\epsilon_n^{(t)} = \frac{C_1}{t+1}$ case is proved.

Case 2: $\epsilon_n^{(t)} = \max_{k\in[N]}\frac{1}{t+1}(a_k^{(t)} + b_{k,n}^{(t)})$.

Next, we analyze the case that $\epsilon_n^{(t)}$ has a faster convergence rate. Combining Claim 9,10 and Eqs. (78) and(79), we show estimations of $A_k(w^{(t)}), B_k(w^{(t)};n)$, $\ell_k(w^{(t)})$ and $\nabla\ell_k(w^{(t)})^\top(\frac{e_n-w^{(t)}}{t+1})$ are accurate enough for us to distinguish $I(w^{(t)};n)$ from $I^c(w^{(t)};n)$, and $I(w^{(t)})$ from $I^c(w^{(t)})$ with a high probability, given a large enough t. Hence, $i\in I(w^{(t)};n)$ with a probability at least $1-\frac{\zeta^{(t)}}{2}$, if t is large enough to satisfy

$$\begin{aligned}&\frac{2(t+2)a_i^{(t)}}{t+1} + \frac{2b_{i,n}^{(t)}}{t+1}\leq L_{SimCos}(w^{(t)};e_n)\\ &- \max_{k\in I^c(w^{(t)};n)}\left[\frac{t+2}{t+1}A_k(w^{(t)}) - B_k(w^{(t)};n) - \frac{\mu_k^2}{(\mu_k - o_k)^2}\right].\end{aligned} \tag{90}$$

And, $j\in I(w^{(t)})$ with a probability at least $1 - \frac{\zeta^{(t)}}{4}$, if t satisfies

$$\begin{aligned}&2a_j^{(t)}\leq L_{SimCos}(w^{(t)})\\ &\quad - \max_{k\in I^c(w^{(t)})}\left[A_k(w^{(t)}) - \frac{\mu_k^2}{(\mu_k - o_k)^2}\right].\end{aligned} \tag{91}$$

To summarize, if t satisfies Claim 11, Eqs.(90) and (91), then with a probability at least $1 - \zeta^{(t)}$ that

$$\begin{aligned}&\left\{L_{SimCos}(w^{(t)};e_n) - L_{SimCos}(w^{(t)})\right\}\\ &\quad - \left\{\widehat{L}_{SimCos}(w^{(t)};e_n) - \widehat{L}_{SimCos}(w^{(t)})\right\}\\ &= \left\{\ell_{i'}(w^{(t)}) + \nabla\ell_{i'}(w^{(t)})^\top\frac{e_n-w^{(t)}}{t+1} - \ell_{j'}(w^{(t)})\right\}\\ &\quad - \left\{\hat{\ell}_i(w^{(t)}) + \hat{\nabla}\ell_i(w^{(t)})^\top\frac{e_n - w^{(t)}}{t+1} - \hat{\ell}_j(w^{(t)})\right\}\\ &\overset{(a)}{=}\left\{\ell_i(w^{(t)}) + \nabla\ell_i(w^{(t)})^\top\frac{e_n-w^{(t)}}{t+1} - \ell_j(w^{(t)})\right\}\\ &\quad - \left\{\hat{\ell}_i(w^{(t)}) + \hat{\nabla}\ell_i(w^{(t)})^\top\frac{e_n - w^{(t)}}{t+1} - \hat{\ell}_j(w^{(t)})\right\}\\ &\overset{(b)}{=}\left\{\ell_i(w^{(t)}) + \nabla\ell_i(w^{(t)})^\top\frac{e_n - w^{(t)}}{t+1} - \ell_i(w^{(t)})\right\}\\ &\quad -\left\{\hat{\ell}_i(w^{(t)}) + \hat{\nabla}\ell_i(w^{(t)})^\top\frac{e_n - w^{(t)}}{t + 1} - \hat{\ell}_i(w^{(t)})\right\}\\ &\overset{(c)}{=}\frac{A_i(w^{(t)}) - B_i(w^{(t)};n)}{t+1} - \frac{\widehat{A}_i(w^{(t)}) - \widehat{B}_i(w^{(t)};n)}{t+1}\leq\frac{a_i^{(t)} + b_{i,n}^{(t)}}{t+1},\end{aligned} \tag{92}$$

which is equivalent to

$$tL_{SimCos}(w^{(t)};e_n) - L_{SimCos}(w^{(t)}) - g_n^{(t)}\leq\frac{a_i^{(t)} + b_{i,n}^{(t)}}{t+1}. \tag{93}$$

Assume $t\geq T_0$ is large enough to satisfy the above conditions. Then, (a) is achieved for $i\in I(w^{(t)};n)$, $j\in I(w^{(t)})$ with probabilities at least $1 - \zeta^{(t)}/2$ and $1 - \zeta^{(t)}/4$; (b) is achieved for $i = j$ with a probability at least $1 - \zeta^{(t)}/4$; and (c) can be shown by plugging $A_i(w^{(t)})$, $B_i(w^{(t)};n)$, $\widehat{A}_k(w^{(t)})$, and $\widehat{B}_k(w^{(t)};n)$. We analyze the selection of T_0 in the proof of Theorem 7.

Similarly, with a probability at least $1 - \zeta^{(t)}$, it holds that

$$\begin{aligned}&L_{SimCos}(w^{(t)};e_n) - L_{SimCos}(w^{(t)}) - g_n^{(t)}\\ &=\frac{A_i(w^{(t)}) - B_i(w^{(t)};n)}{t+1} - \frac{\widehat{A}_i(w^{(t)}) - \widehat{B}_i(w^{(t)};n)}{t+1}\geq - \frac{a_i^{(t)} + b_{i,n}^{(t)}}{t+1}.\end{aligned} \tag{94}$$

Namely, $\epsilon_n^{(t)}\leq\max_{i\in[N]}\frac{1}{t+1}(a_i^{(t)} + b_{i,n}^{(t)})$ when t is large enough to satisfy Claim 11, Eqs. (90) and (91). □

Theorem 6 provides upper bounds of $\epsilon_n^{(t)}$ under different conditions, depending on C_1, $a_i^{(t)}$ and $b_{i,n}^{(t)}$. And we bound C_1 in Theorem 6, and bound $a_i^{(t)}$ and $b_{i,n}^{(t)}$ in Claim 9 and 10.

Step 7: Finally, we give formal regret bounds of the SimCos algorithm and prove Theorem 7.

For convenience, assume that:

$$\begin{aligned}\rho = \min_{w^{(t)},h}&L_{SimCos}(w^{(t)};e_h)\\ &- \max_{k\in I^c(w^{(t)};h)}\ell_k(w^{(t)};e_h) + \nabla\ell_k(w^{(t)})^\top\frac{e_h - w^{(t)}}{t+1},\end{aligned} \tag{95}$$

$$\gamma = \min_{w^{(t)}}L_{SimCos}(w^{(t)}) - \max_{k\in I^c(w^{(t)})}\ell_k(w^{(t)}). \tag{96}$$

Proof of Theorem 7. Specifically, we show that

$$\mathbb{E}R_T \leq 2\left[C_1 T_0 - C_2(\ln\sqrt{8}T_0)^2 - 2C_3\sqrt{T_0 \ln 8T_0^2}\right]\cdot\tfrac{1}{T} + 3(\beta' - \alpha')\cdot\tfrac{\ln T}{T} + 2C_2\tfrac{(\ln\sqrt{8}T)^2}{T} + 4C_3\sqrt{\tfrac{\ln 8T^2}{T}}, \tag{97}$$

where $$C_1 = \max_{k\in[N]}\tfrac{2\xi^2\mu_k}{(\mu_k - o_k)^2} + \max_{k\in[N]}\tfrac{2\xi^3}{(\mu_k - o_k)^2}, \tag{98}$$

$$C_2 = \max_{k\in[N]}\tfrac{2\xi^2(1+\xi)}{3(\mu_k - o_k)^2}, \tag{99}$$

$$C_3 = \max_{k\in[N]}\tfrac{\sqrt{2\xi^3\mu_i}(1+\xi)}{(\mu_i - o_i)^2}, \tag{100}$$

$$C_4 = \max_{i,j\in[N]}\tfrac{(\frac{2}{\xi}+1)(\xi^2-1)\mu_j}{\gamma(\mu_j - o_j)^2} + \tfrac{(1-\frac{1}{\xi})(1+\frac{2}{\xi})\mu_i^2}{\gamma(\mu_i - o_i)^2}, \tag{101}$$

$$C_5 = \min\left\{\tfrac{\rho(C_4+1)(\xi+1)}{2(C_4+\xi+2)}, \tfrac{\gamma}{2+\xi}\right\}, \tag{102}$$

$$T_0 = \max\left\{C_4, \left(\tfrac{C_3 + \sqrt{C_3^2 + 4C_3C_5}}{2C_5}\right)^4, 150\right\}. \tag{103}$$

Let $T_0 - 1$ be the last time before T such that Eqs. (84), (90) and (91) are not all satisfied. Then

$$\begin{aligned}\mathbb{E}\textstyle\sum_{t\in[T]} c_{I_t}^{(t)} &\overset{(a)}{\leq} C_1\left(T_0 - 1 + \frac{2}{T_0}\right) + \textstyle\sum_{t=T_0}^T \max_{k\in[N]} a_k^{(t)} + b_{k,I_t}^{(t)}\\ &\overset{(b)}{\leq} C_1 T_0 + C_2\textstyle\sum_{t=T_0}^T \frac{\ln(8/\zeta^{(t)})}{t} + C_3\textstyle\sum_{t=T_0}^T\sqrt{\frac{\ln(8/\zeta^{(t)})}{t}},\\ &\overset{(c)}{\leq} C_1 T_0 + C_2(\ln\sqrt{8}t)^2\Big|_{T_0}^{T} + 2C_3\sqrt{t\ln 8t^2}\Big|_{T_0}^{T},\\ &= \left[C_1 T_0 - C_2(\ln\sqrt{8}T_0)^2 - 2C_3\sqrt{T_0\ln 8T_0^2}\right]\\ &\quad + C_2(\ln\sqrt{8}T)^2 + 2C_3\sqrt{T\ln 8T^2}.\end{aligned} \tag{104}$$

Notice that (a) is achieved by Theorem 6 and $\sum_t \zeta^{(t)} \leq \frac{2}{T_0}$; (b) is achieved by plugging upper bounds of $a_k^{(t)}$ and $b_{k,I_t}^{(t)}$ in Claim 9 and 10; and (c) is achieved for

$$\textstyle\sum_{t=a}^b \frac{\ln(8/\zeta^{(t)})}{t} < \left(\ln\sqrt{8}t\right)^2\Big|_a^b, \tag{105}$$

$$\begin{aligned}\textstyle\sum_{t=a}^b\sqrt{\frac{\ln(8/\zeta^{(t)})}{t}} &< 2\sqrt{t\ln 8t^2} - 4\sqrt{t}D_+\left(\frac{1}{2}\sqrt{\ln 8t^2}\right)\Big|_a^b\\ &< 2\sqrt{t\ln 8t^2}\Big|_a^b,\end{aligned} \tag{105}$$

where $D_+(\cdot)$ is the Dawson's integral. Combine Eqs. (77) and (104), we have Eq. (97). Next, we bound T_0. By the upper bounds of $a_k^{(t)}, b_{k,h}^{(t)}$ given by Claim 9 and 10, and Eqs. (99) and (100), we have $\forall k\in[N]$ that

$$\frac{2(t+2)a_k^{(t)}}{t+1} + \frac{2b_{k,h}^{(t)}}{t+1} \leq \frac{2(t+\xi+2)}{(t+1)(\xi+1)}\left(C_2\frac{\ln(8/\zeta^{(t)})}{t} + C_3\sqrt{\tfrac{\ln(8/\zeta^{(t)})}{t}}\right).$$

Hence, Eq. (90) is satisfied if

$$\frac{2(t+\xi+2)}{(t+1)(\xi+1)}\left(C_2\frac{\ln(8/\zeta^{(t)})}{t} + C_3\sqrt{\frac{\ln(8/\zeta^{(t)})}{t}}\right) \leq \rho. \tag{107}$$

Similarly, Eq. (91) is satisfied if

$$2\left(C_2\frac{\ln(8/\zeta^{(t)})}{t} + C_3\sqrt{\tfrac{\ln(8/\zeta^{(t)})}{t}}\right) \leq \gamma. \tag{108}$$

With Eq. (96), in the right hand side of Eq. (84), we have

$$\begin{aligned}&\left[A_i(w^{(t)}) - \frac{\mu_i^2}{(\mu_i - o_i)^2}\right] - \left[A_j(w^{(t)}) - \frac{\mu_j^2}{(\mu_j - o_j)^2}\right]\\ &= L_{SimCos}(w^{(t)}) - \ell_j(w^{(t)}) \geq \gamma.\end{aligned} \tag{109}$$

Notice that $A_j(w^{(t)}) \leq \frac{\xi\mu_j}{(\mu_j - o_j)^2}$. By relaxing Eq. (108) to

$$(2+\xi)\left(C_2\frac{\ln(8/\zeta^{(t)})}{t} + C_3\sqrt{\tfrac{\ln(8/\zeta^{(t)})}{t}}\right) \leq \gamma, \tag{110}$$

we have $\forall k$, $(2+\xi)a_k^{(t)} \leq \gamma$. Then the right hand side of Eq. (84) is upper bounded by

$$\begin{aligned}&\frac{1}{\gamma - a_i^{(t)} - a_j^{(t)}}\left[\frac{(\xi^2-1)\mu_j - (1-\frac{1}{\xi})\mu_j^2}{(\mu_j - o_j)^2} + \frac{(1-\frac{1}{\xi})\mu_i^2}{(\mu_i - o_i)^2} + \left(\xi - \frac{1}{\xi}\right)a_j^{(t)}\right]\\ &\quad - 2 + \frac{1}{\xi} \leq \tfrac{(\frac{2}{\xi}+1)(\xi^2-1)\mu_j}{\gamma(\mu_j - o_j)^2} + \tfrac{(1-\frac{1}{\xi})(1+\frac{2}{\xi})\mu_i^2}{\gamma(\mu_i - o_i)^2}\\ &\leq \max_{i,j\in[N]}\tfrac{(\frac{2}{\xi}+1)(\xi^2-1)\mu_j}{\gamma(\mu_j - o_j)^2} + \tfrac{(1-\frac{1}{\xi})(1+\frac{2}{\xi})\mu_i^2}{\gamma(\mu_i - o_i)^2} \triangleq C_4.\end{aligned} \tag{111}$$

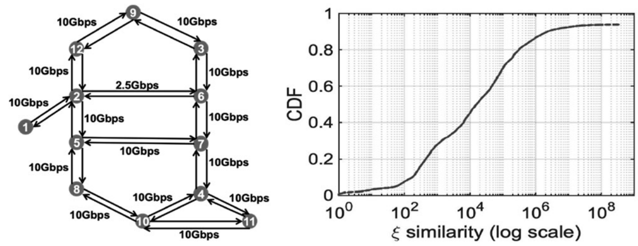
Hence, it is sufficient to say Eq. (84) is achieved if $t \geq C_4$. As $\forall t \geq 150$, $\frac{\ln 8t^2}{t} \leq \frac{1}{\sqrt{t}}$. By solving $C_2 t^{-\frac{1}{2}} + C_3 t^{-\frac{1}{4}} = C_5$

$$t = \left(\frac{C_3 + \sqrt{C_3^2 + 4C_3C_5}}{2C_5}\right)^4, \tag{112}$$

which guarantees that both Eqs. (107) and (108) can be satisfied. Hence, T_0 is upper bounded by Eq. (103). □

6 Applications

In this section, we demonstrate the versatility of MIS-Learning framework by applying it to evaluate the risks for a set of rare threats in two applications. In the first application, we consider the Internet backbone networks. We evaluate the impact of *net- work link failures* on the occurrences of interested events $\mathcal{E}_n$, which are specified as the *non-satisfaction of bandwidth demands for traffic flows* n. In the second application, we consider smart grids. We study the impact of *network component failures* on the cascading failures $\mathcal{E}_n$ of



(a) The Abilene network. (b) CDF of ξ similarity.

Fig. 2. The backbone network topology and ξ-similarity information.

transmission lines n and $n \in [N]$. Numerical results show that compared with the uniform mixture IS, our SumVar and SimCos algorithms reduce the associated cost measures by 37.8 and 61.6 percent in the backbone network application, and by 56.4 and 68.8 percent in the smart grid application.

6.1 Application to Backbone Network

Problem Description. We consider the Abilene backbone network [18], [19]. As depicted in Fig. 2a, the network contains 12 nodes and 30 links. Each link fails with a probability of 0.01 and x denotes the failure occurrence profile. The topology and traffic matrices are collected from [20]. There are 132 competing flows, and their bandwidth demands are extracted from [9]. The flow routing follows the shortest path policy. The capacity allocation follows the max-min fairness policy, which is also adopted by Google's B4 backbone network [21]. We want to evaluate the risks of *link failures* (indicated by x) and study their impact on the occurrence probabilities of *flow demand unsatisfactions*(indicated by $\mathcal{E}_n$), utilizing our MIS-Learning framework.

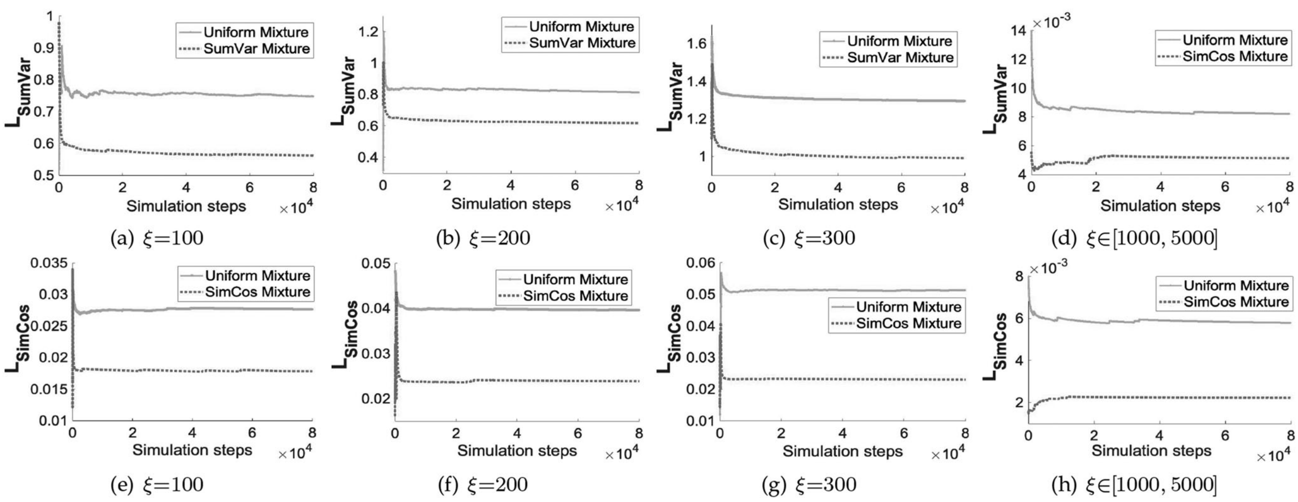
For each interested event $\mathcal{E}_n$ with occurrence probability μ_n, we take the *customized pure IS distribution in* [9] as the efficient IS distribution $Q_n(x)$ of $\mathcal{E}_n$. To accurately estimate $\{\mu_n\}$ for a set of events $\{\mathcal{E}_n\}$, authors in [9] consider the MIS solution with a *uniform mixture* $w = (\frac{1}{N}, \ldots, \frac{1}{N})$. In the following, we apply our MIS-Learning framework to learn a more efficient mixture w^* which minimizes the cost measure $L(\boldsymbol{\sigma}(w))$.

We first derive ξ-similarities between any two interested events $\mathcal{E}_{n_1}$ and $\mathcal{E}_{n_2}$, $n_1, n_2 \in [N]$. The cumulative probability distribution (CDF) of the *pairwise* ξ-similarity is provided in Fig. 2b. By setting upper thresholds of the pairwise ξ-similarity, we can partition $\{\mathcal{E}_n\}_{n=1}^{N}$ into different subsets, on which we apply our MIS-Learning method to find an efficient w to estimate occurrence probabilities of events simultaneously. We set the upper bounds of the pairwise ξ-similarity as $\xi \leq 100$, $\xi \leq 200$, $\xi \leq 300$ and $\xi \in [1000, 5000]$, and obtain corresponding event subsets $\{\mathcal{E}_n\}_{n=1}^{N'}$ with set sizes of $N' = 16$, $N' = 19$, $N' = 30$ and $N' = 5$.

Minimizing the Sum of Variances. We start with the SumVar MIS-Learning with $L(\boldsymbol{\sigma}(w)) \triangleq L_{\text{SumVar}}(w)$. For each event subset $\{\mathcal{E}_n\}_{n=1}^{N'}$ with the corresponding ξ-similarity threshold, we run the SumVar MIS-learning for 80,000 rounds. Fig. 3 plots the cost measure $L_{\text{SumVar}}(w)$ in each round. We then compare the result with the uniform mixture proposed in [9]. Figs. 3a, 3b and 3c illustrate the reduction of $L_{\text{SumVar}}(w)$ achieved by the SumVar MIS-Learning with a small ξ-similarity. Fig. 3d illustrates the performance of the SumVar MIS-Learning with a large ξ-similarity. The SumVar MIS-Learning with Algorithm 1 reduces the cost measure by 25.1, 23.6, 26.4 and 37.8 percent when $\xi \leq 100$, $\xi \leq 200$, $\xi \leq 300$ and $\xi \in [1000, 5000]$.

Minimizing the Simulation Cost. We then consider the SimCos MIS-Learning with $L(\boldsymbol{\sigma}(w)) \triangleq L_{\text{SimCos}}(w)$. For each event subset $\{\mathcal{E}_n\}_{n=1}^{N'}$ with the corresponding ξ-similarity threshold, we run the SimCos MIS-Learning for 80,000 round. Fig. 3 plots $L_{\text{SimCos}}(w)$ in each round. Figs. 3e, 3f and 3g show the reduction of $L_{\text{SimCos}}(w)$ achieved by the SimCos MIS-Learning with a small ξ-similarity, while Fig. 3h show the reduction with a large ξ-similarity. The SimCos MIS-Learning reduces the cost measure by 35.7, 55.1, 39.9 and 61.6 percent when $\xi \leq 100$, $\xi \leq 200$, $\xi \leq 300$ and $\xi \in [1000, 5000]$.

Impact of ξ-Similarity on the Convergence Rate. We study the convergence rate of cost measures in Fig. 4, and compare convergence rates under large ξ (i.e., $\xi \in [1000, 5000)$ and small ξ (i.e., $\xi \leq 300$). For the SumVar MIS-Learning with Algorithm 1, Theorem 4 implies that the regret $L_{SumVar}(w) - L_{SumVar}(w^*)$ first decreases at a fast rate in Eq. (22) and then at a slow rate in Eq. (23). Theorem 4 also reveals that a



(a) ξ=100 (b) ξ=200 (c) ξ=300 (d) $\xi \in [1000, 5000]$

(e) ξ=100 (f) ξ=200 (g) ξ=300 (h) $\xi \in [1000, 5000]$

Fig. 3. The reduction of cost measure $L_{SumVar}(x)$ (or $L_{SimCos}(x)$) achieved by MIS-learning, compared with the uniform mixture. (a)-(d) show the SumVar case and (e)-(h) show the SimCos case; (a)-(c), (e)-(g) show the small ξ case, and (d), (f) show the large ξ case.

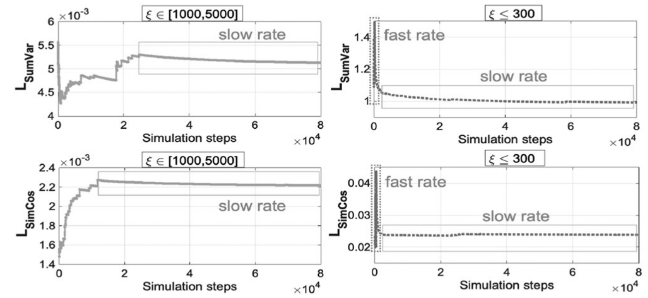
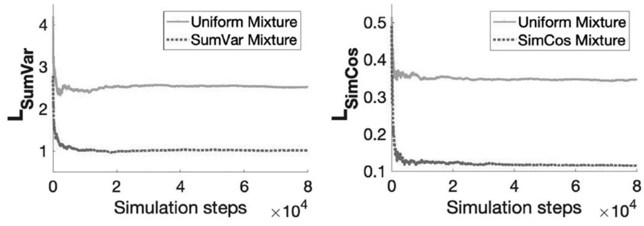


Fig. 4. The impact of ξ-similarity on the convergence rate.



(a) SumVar MIS-Learning. (b) SimCos MIS-Learning.

Fig. 6. The reduction of cost measure $L_{SumVar}(x)$ (or $L_{SimCos}(x)$) achieved by MIS-Learning, in smart grid cascading failure simulation.

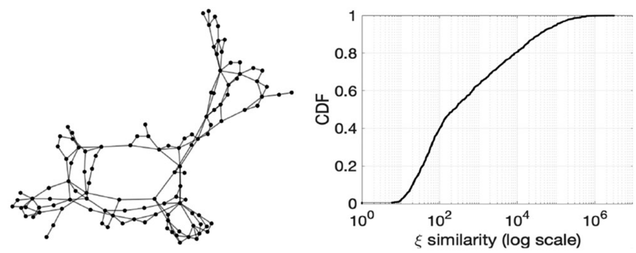
smaller ξ implies a longer fast rate period. As shown in Fig. 4, with a small ξ, $L_{SumVar}(w)$ decreases first at a fast rate and then at a slow rate; with a large ξ, the fast rate period vanishes. For SimCos MIS-Learning with Algorithm 2, Theorem 7 states that the regret decreases first at a fast rate of $O(1/T)$ and then at a slow rate of $Q(\sqrt{\ln T/T})$ in Eq. (97). As shown in Fig. 4, with a small ξ, $L_{\text{SimCos}}(w)$ decreases first at a fast rate and then at a slow rate; with a large ξ, the short fast rate period vanishes.

6.2 Application to Smart Grid

Problem Description. We consider a realistic smart grid, i.e., the IEEE 118-bus test system containing 118 buses, 177 transmission lines and 9 generators [22]. It can be simplified as a network with 118 nodes and 186 links [23], as illustrated in Fig. 5a. Each link fails with a probability uniformly selected from $0.01\cdot[95\%, 105\%]$, and the occurrences of initial link failures are indicated by x.

Assume the power flows on each link follow the nonlinear AC model [23]. When an initial set of links fail, the smart grid can be divided into one or more connected components, each of which can operate autonomously. Components with no supply or no demand become dead, and the supply and demand within alive components are rebalanced. This rebalance changes the power flows on each link and result in the outages of more links, i.e., cascading failures [23]. Assume the cascade process follows the shedding and curtailing balancing rule [23], [24], and deterministic link outage rule [23]. We aim to evaluate the risks of *link initial failures* (indicated by x), and study their impact on the occurrence probabilities of *link cascading failures* (indicated by $\mathcal{E}_n$), utilizing our MIS-Learning framework.

Similar with the backbone network scenario, for each $\mathcal{E}_n$ with occurrence probability μ_n, we take the *customized pure IS distribution in* [9] as $Q_n(x)$ of $\mathcal{E}_n$. We derive ξ-similarities



(a) IEEE 118-bus power system. (b) CDF of ξ similarity.

Fig. 5. The smart grid topology and ξ-similarity information.

for any $\mathcal{E}_{n_1}$ and $\mathcal{E}_{n_2}$, $n_1, n_2 \in [N]$, and plot the CDF of such pairwise ξ-similarity in Fig. 5b. We set the upper bound of ξ as $\xi \leq 100$ and obtain an event subset $\{\mathcal{E}_n\}_{n=1}^{N'}$ with $N' = 30$. We apply our MIS-Learning framework to learn an efficient mixture w^* minimizing the cost measure $L(\boldsymbol{\sigma}(w))$ of $\{\mathcal{E}_n\}_{n=1}^{N'}$. The cost measure reduction achieved by w^* is compared with the reduction achieved by the *uniform mixture* $w = (\frac{1}{N'}, \ldots, \frac{1}{N'})$ in [9].

Minimizing the Sum of Variances. We start with the SumVar MIS-Learning with $L(\boldsymbol{\sigma}(w)) \triangleq L_{\text{SumVar}}(w)$. We run the SumVar MIS-Learning for 80,000 rounds. Fig. 6a plots the cost measure $L_{\text{SumVar}}(w)$ in each round. One can observe that the SumVar MIS-Learning with Algorithm 1 reduces the cost measure by 56.4 percent compared with the uniform mixture.

Minimizing the Simulation Cost. We consider the SimCos MIS-Learning with $L(\boldsymbol{\sigma}(w)) \triangleq L_{\text{SimCos}}(w)$. We run the SumVar MIS-Learning for 80,000 round. Fig. 6b plots the cost measure $L_{\text{SimCos}}(w)$ in each round. One can observe that the SimCos MIS-Learning with Algorithm 2 reduces the cost measure by 68.8 percent compared with the uniform mixture.

7 Related Work

7.1 MIS-Learning versus IS and MIS

Comprehensive reviews on the rare event simulation are given in [25], [26]. These works are mainly IS based and focus on single rare event estimation [27]. Given many rare events to estimate, as each $Q_n(x)$ is merely *customized* for $\mathcal{E}_n$ and may not work efficiently for other events, IS needs to *"sequentially"* estimate the occurrence of each $\mathcal{E}_n$ with its corresponding pure importance distribution $Q_n(x)$.

To efficiently estimate multiple rare events, various works [9], [28], [29] consider using the MIS to cooperate multiple $Q_n(x)$. However, most MIS based works take a *uniform mixture* [9] or *heuristic mixture strategies without theoretical guarantees* [30]. Authors in [28], [29] provide examples on the impact of mixture on the estimation efficiency and intuitive guidelines for selecting a proper mixture. Some works [31], [32] consider computing the optimal mixture via standard convex optimization methods. However, they require that at each iteration, the variances (i.e., their cost measure) should either *be computed analytically* [32] or *be estimated accurately from sufficient samples* [31], which is unrealizable or computational expensive for the curse of dimensionality.

Our work aims to efficiently learn the optimal mixture working for estimations of many rare events, with a zero cost on extra samples. We reveal that *not all rare events can be efficiently estimated at the same time*, and we introduce the

ξ-similarity to partition events into subsets with smaller ξ values, which can be efficiently estimated via MIS at the same time.

Note that the IS distribution design can also jointly consider multiple rare events $\{\mathcal{E}_n\}_{n=1}^N$. However, both the objective for designing such an IS distribution and the methods to optimize this objective need a careful design. And this is quite different from designing IS for a single rare event, where we only need to consider minimizing the variance for estimating $\mathcal{E}_n$. As the basic idea of IS is to use a distribution $Q(x)$ rather than the original distribution $P(x)$ to generate x samples. The mixture distribution $Q(x;w) = \sum_{n\in[N]} w_n Q_n(x)$ obtained using our method can be considered as an IS distribution $Q(x)$ designed for $\{\mathcal{E}_n\}_{n=1}^N$.

7.2 MIS-Learning versus Stochastic Optimization

The MIS-Learning can be viewed as *the stochastic optimization (SO) problem over the simplex*: to minimize the objective function $L(\boldsymbol{\sigma}(w))$, we choose at each round an action I_t, which affects the variable w and provides observations on $L(\boldsymbol{\sigma}(w))$.

In the common case where objectives are *smooth*, i.e., $L(\boldsymbol{\sigma}(w)) \triangleq L_{SumVar}(w)$, *iterative gradient-based methods*, such as the gradient descent (GD) and stochastic gradient descent (SGD) [33], are popular optimization tools. Yet in our setting, *neither the gradient $\nabla L_{SumVar}(w)$ nor its components can be computed exactly* and so estimations are required. To accurately estimate $\nabla L_{SumVar}(w^{(t)})$ and meanwhile guarantee a good convergence speed, SGD needs to generate sufficient simulation samples from $Q(x;w^{(t)})$ at each learning round t, making the learning cost unaffordable.

When objectives are *non-smooth*, i.e., a pointwise maximum function $L(\boldsymbol{\sigma}(w)) \triangleq L_{SimCos}(w)$ with smooth components, *gradient mapping based methods* [16] guarantee an exponential regret convergence. Yet in our setting, it faces the same problem of *expensive gradient (or its components) estimation*. A more challenging point is *the constrained $w^{(t)}$ updating*: the updating of $w^{(t)}$ has a fixed step size of $1/t$ and constrained moving directions, i.e., $w^{(t)} = w^{(t-1)} + \frac{1}{t}(e_{I_t} - w^{(t-1)})$.

Our method solves these challenges and reduce the gradient (or its components) estimation cost by generating only one sample x from one of $\{Q_n(x)\}_{n=1}^N$ at each round. Hence, it has a "*zero cost on extra samples*". Besides, with SO, estimations of rare events $\{\mathcal{E}_n\}_{n=1}^N$ are performed only after deriving a proper w. In other words, samples generated while optimizing w cannot be used for estimating $\{\mathcal{E}_n\}_{n=1}^N$. As a contrast, our method estimates $\{\mathcal{E}_n\}_{n=1}^N$ and learns the optimal mixture w^* at the same time. Thus, it also has a "*zero learning cost*".

7.3 MIS-Learning versus MAB Optimization

The MIS-Learning is also similar to the MAB optimization [34], [35], where at each round t, we pick an action e_{I_t} and observe information on the loss function L. The major difference is that these works consider a *cumulative regret* $\frac{1}{T}\sum_{t\in[T]} L(e_{I_t})$ but we focus on the *global loss* $L(\frac{1}{T}\sum_{t\in[T]} e_{I_t})$.

Problems related to the MAB optimization with the global loss have been studied in [14], [36], [37], [38], where they consider minimizing a known loss $L(w^{(t)\top}V)$ with an unknown matrix V. This differs from our setting where L is unknown and cannot be computed analytically. [36], [38] consider a stochastic setting and achieve a convergence rate of $O(\sqrt{1/T})$. The work in [37] considers an adversarial setting, but there are cases that their regrets cannot converge to zero. Our SumVar case is similar to [14], which considers the global loss $L(\frac{1}{T}\sum_{t\in[T]} e_{I_t})$ and focuses on the strongly-convex and smooth loss function L. They consider $L(w) \triangleq \sum_{n\in[N]} \sigma_n^2/w_n$ with the unknown but fixed σ_n^2, $n\in[N]$. Yet, in our setting, $\sigma_n^2, n\in[N]$ also depend on w.

8 Conclusion

This paper aims at providing efficient risk evaluation on many network rare threats. We develop an MAB OL framework to address the high simulation cost limitation of IS in estimating occurrence probabilities for *a set of rare threats*. Our framework consists of a mixture importance sampling optimization problem (MISO) and two OL algorithms. MISO aims to select the optimal mixture w^* attaining various tradeoffs, which are quantified by two cost measures. We first show the objective function of MISO is computationally expensive to evaluate. Then we extend MISO to an OL setting to efficiently optimize the objective function without incurring any extra learning cost. Our SumVar and SimCos algorithms learn to minimize the *sum of variances* and *simulation cost* with regrets of $(\ln T)^2/T$ and $\sqrt{\ln T/T}$ respectively, where T is the number of samples. We demonstrate our method on various realistic applications, and our method reduces the cost measure value by as high as 61.6 percent in the backbone network scenario, and by 68.8 percent in the smart grid scenario.

Acknowledgments

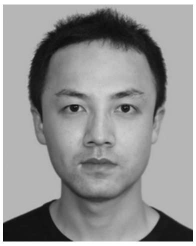
The work of Hong Xie was supported by Chongqing NSF (cstc2020jcyj-msxmX0652) and Fundamental Research Funds for Central Universities (2020CDJ-LHZZ-057). The work of John C.S. Lui was supported in part by the GRF 14200420.

References

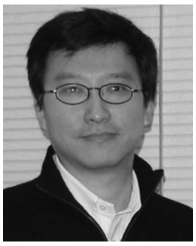
[1] L. Mikeev, W. Sandmann, and V. Wolf, "Efficient calculation of rare event probabilities in Markovian queueing networks," in *Proc. 5th Int. ICST Conf. Perform. Eval. Methodologies Tools*, 2011, pp. 186–196.
[2] F. Cadini, G. L. Agliardi, and E. Zio, "Estimation of rare event probabilities in power transmission networks subject to cascading failures," *Rel. Eng. Syst. Saf.*, vol. 158, pp. 9–20, 2017.
[3] H. Cancela, M. E. Khadiri, G. Rubino, and B. Tuffin, "Balanced and approximate zero-variance recursive estimators for the network reliability problem," *ACM Trans. Model. Comput. Simul.*, vol. 25, no. 1, pp. 1–19, 2014.
[4] H. Cancela and M. ElKhadiri , "The recursive variance-reduction simulation algorithm for network reliability evaluation," *IEEE Trans. Rel.*, vol. 52, no. 2, pp. 207–212, Jun. 2003.
[5] T. Elperin, I. Gertsbakh, and M. Lomonosov, "Estimation of network reliability using graph evolution models," *IEEE Trans. Rel.*, vol. 40, no. 5, pp. 572–581, Dec. 1991.
[6] G. S. Fishman, "A Monte Carlo sampling plan for estimating network reliability," *Operations Res.*, vol. 34, no. 4, pp. 581–594, 1986.
[7] I. B. Gertsbakh *et al.*, *Statistical Reliability Theory*. Boca Raton, FL, USA: CRC Press, 1988.
[8] G. Rubino *et al.*, *Rare Event Simulation Using Monte Carlo Methods*. Hoboken, NJ, USA: Wiley, 2009.
[9] T. Liu and J. C. S. Lui, "FAVE: A fast and efficient network flow availability estimation method with bounded relative error," in *Proc. IEEE Conf. Comput. Commun.*, 2019, pp. 1783–1791.

[10] Z. I. Botev, P. L'Écuyer, G. Rubino, R. Simard, and B. Tuffin, "Static network reliability estimation via generalized splitting," *INFORMS J. Comput.*, vol. 25, no. 1, pp. 56–71, 2013.
[11] Z. I. Botev, P. L'Ecuyer, and B. Tuffin, "Dependent failures in highly reliable static networks," in *Proc. Winter Simul. Conf.*, 2012, pp. 1–12.
[12] P. W. Glynn and D. L. Iglehart, "Importance sampling for stochastic simulations," *Manag. Sci.*, vol. 35, pp. 1367–1392, 1989.
[13] A. Rai, R. C. Valenzuela, B. Tuffin, G. Rubino, and P. Dersin, "Approximate zero-variance importance sampling for static network reliability estimation with node failures and application to rail systems," in *Proc. IEEE Winter Simul. Conf.*, 2016, pp. 3201–3212.
[14] Q. Berthet and V. Perchet, "Fast rates for bandit optimization with upper-confidence frank-wolfe," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 2225–2234.
[15] M. Frank and P. Wolfe, "An algorithm for quadratic programming," *Naval Res. Logistics Quart.*, vol. 3, no. 1/2, pp. 95–110, 1956.
[16] Y. Nesterov, "Introductory lectures on convex optimization: a basic course," Springer, 1st ed. 2003.
[17] S. Bernstein, "The theory of probabilities," Gastehizdat Publishing House, Moscow, 1946.
[18] Abilene. [Online]. Available: https://www.internet2.edu/
[19] W. Jiang, R. Zhang-Shen, J. Rexford, and M. Chiang, "Cooperative content distribution and traffic engineering in an ISP network," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 37, no. 1, pp. 239–250, 2009.
[20] Y. Zhang, "Six months of Abilene traffic matrices," 2014. [Online]. Available: http://www.cs.utexas.edu/ yzhang/
[21] S. Jain *et al.*, "B4: Experience with a globally-deployed software defined WAN," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 3–14, 2013.
[22] IEEE 118-bus system. [Online]. Available: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/ieee-118-bus-system/
[23] H. Cetinay, S. Soltan, F. A. Kuipers, G. Zussman, and P. Van Mieghem , "Comparing the effects of failures in power grids under the AC and DC power flow models," *IEEE Trans. Netw. Sci. Eng.*, vol. 5, no. 4, pp. 301–312, Fourth Quarter 2018.
[24] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman, "Power grid vulnerability to geographically correlated failures—Analysis and control implications," in *Proc. IEEE Conf. Comput. Commun.*, 2014, pp. 2634–2642.
[25] J. Bucklew, *Introduction to Rare Event Simulation*. Berlin, Germany: Springer, 2013.
[26] G. Rubino and B. Tuffin, *Rare Event Simulation Using Monte Carlo Methods*. Hoboken, NJ, USA: Wiley, 2009.
[27] J. Frank, S. Mannor, and D. Precup, "Reinforcement learning in the presence of rare events," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 336–343.
[28] T. Hesterberg, "Advances in importance sampling," PhD thesis, Stanford Univ., 1988.
[29] A. Owen and Y. Zhou, "Safe and effective importance sampling," *J. Amer. Statist. Assoc.*, vol. 95, no. 449, pp. 135–143, 2000.
[30] R. Kanj, R. Joshi, and S. Nassif, "Mixture importance sampling and its application to the analysis of SRAM designs in the presence of rare failure events," in *Proc. 43rd ACM/IEEE Des. Autom. Conf.*, 2006, pp. 69–72.
[31] J.-M. Cornuet, J.-M. Marin, A. Mira, and C. P. Robert, "Adaptive multiple importance sampling," *Scand. J. Statist.*, vol. 39, pp. 798–812, 2012.
[32] H. Y. He and A. B. Owen, "Optimal mixture weights in multiple importance sampling," 2014, *arXiv:1411.3954*.
[33] G. Papamakarios, "Comparison of modern stochastic optimization algorithms," Edinbuegh Univ. Press, 2014.
[34] E. Hazan, "The convex optimization approach to regret minimization," *Optim. Mach. Learn.*, p. 287303, 2012.
[35] S. Shalev-Shwartz *et al.*, "Online learning and online convex optimization," *Found. Trends Mach. Learn.*, vol. 4, no. 2, pp. 107–194, 2012.
[36] S. Agrawal, N. R. Devanur, and L. Li, "An efficient algorithm for contextual bandits with knapsacks, and an extension to concave objectives," in *Proc. Conf. Learn. Theory*, 2016, pp. 4–18.
[37] A. Rakhlin, K. Sridharan, and A. Tewari, "Online learning: Beyond regret," in *Proc. 24th Annu. Conf. Learn. Theory*, 2011, pp. 559–594.
[38] S. Agrawal and N. R. Devanur, "Bandits with concave rewards and convex knapsacks," in *Proc. 15th ACM Conf. Econ. Comput.*, 2014, pp. 989–1006.

Tingwei Liu received the BE and ME degrees in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2013 and 2016, respectively, and the PhD degree from the Department of Computer Science and Engineering, Chinese University of Hong Kong, Hong Kong. Her research interests include networking algorithm and protocol design, network availability measurement, and online learning algorithms.



Hong Xie (Member, IEEE) received the BEng degree from the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China, in 2010, and the PhD degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, in 2015, under the supervision of Prof. John. C. S. Lui. He is currently a research professor at the College of Computer Science, Chongqing University. His research interests include online learning, networking, and data science. He is a member of the ACM.



John C. S. Lui (Fellow, IEEE) received the PhD degree in computer science from the University of California at Los Angeles, Los Angeles, California. He is currently the Choh-Ming Li chair professor at the Department of Computer Science and Engineering, The Chinese University of Hong Kong. His current research interests include machine learning, online learning (e.g., multi-armed bandit, reinforcement learning), network science, future Internet architectures and protocols, network economics, and network/system security, large scale storage systems. He is an elected member of the IFIP WG 7.3, fellow of ACM, senior research fellow of the Croucher Foundation and was the past chair of the ACM SIGMETRICS (2011-2015). He received various departmental teaching awards and the CUHK Vice-Chancellor's Exemplary Teaching Award. He is a co-recipient of the best paper award in the IFIP WG 7.3 Performance 2005, IEEE/IFIP NOMS 2006, SIMPLEX 2013, and ACM RecSys 2017.