Importance Sampling with Online Learning: Methodology and Applications

Abstract

Importance sampling (IS) is widely used in rare event simulation, but it is costly to deal with *many rare events* simultaneously, i.e., IS has to simulate each rare event with its customized importance distribution one by one. To reduce such cost, we aim to find an efficient mixture importance distribution for multiple rare events, and formulate a mixture importance sampling optimization problem (MISOP) to select the optimal mixture. We show that the "*search direction*" of the mixture is computationally expensive to evaluate, making it challenging to locate the optimal mixture. We then formulate a "*zero learning cost*" online learning framework to estimate the "*search direction*" and learn the optimal mixture from simulation samples of events. To capture the simulation cost budget, we only allow to generate one simulation sample in each round of learning (i.e., zero cost on extra samples). This makes it challenging to estimate the "*search direction*" as well as learn the optimal mixture. We develop two online learning algorithms to address this challenge: (1) learning to minimize the sum of estimation variances with regret of $(\ln T)^2/T$; (2) learning to minimize the simulation cost with regret of $\sqrt{\ln T/T}$, where T denotes the number of samples. We demonstrate our methods on a realistic network and our methods reduce the cost measure value by 61.6%, compared with the uniform mixture IS.

1. Introduction

Rare events refer to events that occur rarely but have catastrophic impacts or consequences. For example, in stock markets, some sudden unforeseen events can result in turmoil and loss of money. In transportation systems, some unexpected failures can result in aircraft and car accidents. To quantify such *rare threats*, one needs to accurately evaluate their risks. The following example illustrates the computational challenge in evaluating such risks.

Example 1. *Consider a large-scaled network. We aim to evaluate the occurrence of a rare threat $\mathcal{E}$, i.e., the failure to provide the promised quality-of-service (QoS) guarantee for a critical flow. The rare threat is induced by a subset of potential causes, e.g., the link and node failures, which are indexed by m, $m\in[M]$ and may occur rarely. Let $\boldsymbol{x}\in\{0,1\}^M$ be the occurrence profile of these causes, which happens with probability $P(\boldsymbol{x})$. $\mathcal{E}$ is represented by a set of $\boldsymbol{x}$, which is usually of a large cardinality, say $O(2^M)$. Then, to evaluate the occurrence probability of $\mathcal{E}$, one needs to do $O(2^M)$ enumerations, implying a high computational complexity.*

Monte Carlo (MC) sampling is a typical method to address the problem illustrated in Example 1. However, to obtain accurate estimations, MC needs to simulate a large number of samples to capture sufficient occurrences of the rare threat. Importance Sampling (IS) method is more efficient than MC in estimating the occurrence of a rare threat, through a customized importance distribution $Q(\boldsymbol{x})$ to "boost" the occurrence of rare threat $\mathcal{E}$. One limitation of IS is that it has to simulate each rare threat with its customized importance distribution one by one. This leads to a high simulation cost when dealing with *a set of rare threats*, for instance:

Example 2. *Consider a set of rare threats denoted by $\{\mathcal{E}_n\}_{n=1}^N$ associated with customized importance distributions $\{Q_n(\boldsymbol{x})\}_{n=1}^N$. For example, the $\mathcal{E}_n$ can be the failure to provide the promised QoS guarantees for the critical flow n. Suppose, to estimate $\mathcal{E}_n$ we need a number of T samples from $Q_n(\boldsymbol{x})$. To estimate all events, we need TN samples. The simulation cost for one sample is not cheap, and when the network size is large, this becomes very expensive.*

To relieve the simulation cost burden of IS, we propose the *mixture importance sampling (MIS)*:

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

Through this, each sample $\boldsymbol{x}$ from $Q(\boldsymbol{x};\boldsymbol{w})$ can be used for all rare threats $\{\mathcal{E}_n\}_{n=1}^N$. We aim to answer two questions: (1) *How to quantify the "simulation cost" for each mixture $\boldsymbol{w}$?* (2) *How to locate the optimal mixture?* To design appropriate simulation cost metrics for $\boldsymbol{w}$, one needs to manage the simulation cost resulted from $Q(\boldsymbol{x};\boldsymbol{w})$ for each rare threat $\mathcal{E}_n$. Such metrics (i.e., cost measure) are functions of $Q(\boldsymbol{x};\boldsymbol{w})$. To search the mixture minimizing the simulation cost, one needs to marginalize $\boldsymbol{x}$ in the metric. However, the sample space of $\boldsymbol{x}$ is 2^M, implying a high computational complexity in evaluating the search direction. To address this challenge, we formulate an online learning framework to estimate the "*search direction*" and learn the optimal mixture from simulation samples of events, i.e., $\boldsymbol{x}$ generated from $\{Q_n(\boldsymbol{x})\}_{n=1}^N$. To manage the simulation cost, we can only generate one sample $\boldsymbol{x}$ from one of $\{Q_n(\boldsymbol{x})\}_{n=1}^N$ in each round, i.e., *zero cost on extra samples*. However, this makes it challenging to estimate the "*search direction*" as well as to learn the optimal mixture. We address these challenges and our contributions are:

- We formulate metrics to quantify the simulation cost for the mixture of importance functions. We formulate a mixture importance sampling optimization problem (MISOP) to select the optimal mixture. We show that the "*search direction*" of mixture is computationally expensive to evaluate, making it challenging to locate the optimal mixture.
- We then formulate an online learning framework to estimate the "*search direction*" and learn the optimal mixture from simulation samples of events. Our learning framework manages the simulation cost by generating one simulation sample $\boldsymbol{x}$ from one of $\{Q_n(\boldsymbol{x})\}_{n=1}^N$ in each round of learning (in other words, zero cost on extra samples).
- We develop two learning algorithms to address such challenges: (1) learning to minimize the sum of variances with regret of $(\ln T)^2/T$; (2) learning to minimize the simulation cost with regret of $\sqrt{\ln T/T}$, where T is the number of samples. For each algorithm, we provide: (1) convexity and smoothness analysis; (2) algorithm to estimate the search direction of $\boldsymbol{w}$ with zero cost on extra samples and provable concentration; (3) regret analysis of the algorithm and revealing the impact of key factors, e.g., similarity of $\{Q_n(\boldsymbol{x})\}_{n=1}^N$, on the regret.
- We demonstrate the efficiency of our methods on a realistic network: our SumVar and SimCos MIS-Learning methods reduce the cost measure value by 37.8% and 61.6% compared with the uniform mixture IS.

2. Problem Formulation

We first present the mixture importance sampling model. Then, we formulate a general framework to optimize and estimate the optimal mixture. Finally, we present two important instances of the framework with different cost measures.

2.1. Mixture Importance Sampling

We consider N interested rare events and we aim to estimate the occurrence probability for each individual event. Each event is induced by a subset of M potential *causes* denoted by $[M]$. Let the indicator variable $x_m \in \{0,1\}$ denote the occurrence of cause $m \in [M]$, where $x_m=1$ (or 0) implies the cause m occurs (or not). We denote the *occurrence profile* for all the M causes as the vector $\boldsymbol{x}=(x_1,...,x_M)\in\Omega$, where $\Omega \triangleq \{0,1\}^M$. We formally denote the event $n\in[N]$ as $\mathcal{E}_n \subset \Omega$. Let $P(\boldsymbol{x})\in[0,1]$ be the probabilities for the profile $\boldsymbol{x}$, where $\sum_{\boldsymbol{x}\in\Omega} P(\boldsymbol{x})=1$. Let the indicator function $\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})$ indicate the occurrence of event $\mathcal{E}_n$:

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

The occurrence probability of $\mathcal{E}_n$ is:

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

In many real life applications, the exact value of μ_n is computationally expensive to evaluate, due to a large cardinality of $\mathcal{E}_n$. For instance, consider an Internet-scale network with M physical links, of which the failure of m-th link happens with a probability p_m. There are N competing flows, of which the undelivery of the n-th flow is represented by $\mathcal{E}_n$. We have $P(\boldsymbol{x}) = \prod_{m\in[M]} p_m^{x_m}(1-p_m)^{1-x_m}$. Due to the high complexity of traffic engineering, $\mathcal{E}_n \subset \Omega$ 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.

The rare occurrence of $\mathcal{E}_n$ makes it costly to estimate μ_n by simulating $\boldsymbol{x}$ with $P(\boldsymbol{x})$, i.e., the classical MC method. One typical method to address this challenge is the IS method. Assume for each event $\mathcal{E}_n$, we have a *customised* pure importance distribution $Q_n(\boldsymbol{x})$. IS provides efficient estimation of μ_n if taking $Q_n(\boldsymbol{x})$ to simulate $\boldsymbol{x}$, yet may not work for other events. The *one-run variance* for estimating μ_n with $Q_n(\boldsymbol{x})$ to simulate $\boldsymbol{x}$ is:

$$\mathbb{V}_{\boldsymbol{x}\sim Q_n}\left[\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\tfrac{P(\boldsymbol{x})}{Q_n(\boldsymbol{x})}\right] \triangleq \mathbb{E}_{\boldsymbol{x}\sim Q_n}\left[\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\tfrac{P^2(\boldsymbol{x})}{Q_n^2(\boldsymbol{x})}\right] - \mu_n^2, \tag{3}$$

which determines the simulation cost. Here $\{Q_n(\boldsymbol{x})\}_{n=1}^N$ can be obtained from the IS and Sequential IS methods in (Liu & Lui, 2019).

Yet, given limited simulation budgets and a large N, we could not afford to estimate each μ_n with the corresponding $Q_n(\boldsymbol{x})$ "sequentially": What we need is an efficient sampling distribution which works for *multiple* interested events at the same time. Assume we take a mixture distribution of $\{Q_n(\boldsymbol{x})\}_{n=1}^N$. Formally, we have

$$Q(\boldsymbol{x};\boldsymbol{w}) \triangleq \textstyle\sum_{n\in[N]} w_n Q_n(\boldsymbol{x}), \tag{4}$$

where $\boldsymbol{w} \triangleq (w_1,\ldots,w_N)$, $w_n \geq 0$ and $\sum_{n\in[N]} w_n = 1$. For the ease of presentation, denote the set of all possible choices of $\boldsymbol{w}$ as the probability simplex $\Delta \triangleq \{\boldsymbol{w} | w_n \geq 0, \sum_{n=1}^N w_n = 1\}$.

We define "ξ-*similarity*" to quantify how well the occurrences of interested events $\{\mathcal{E}_i\}_{i=1}^N$ can be efficiently estimated together.

Definition 1 (ξ**-similarity).** *Events* $\{\mathcal{E}_n\}_{n=1}^N$ *are* ξ*-similar, where* $\xi\in[0,\infty]$, *if their corresponding pure importance distributions* $\{Q_n(\boldsymbol{x})\}_{n=1}^N$ *satisfy:*

$$\forall \boldsymbol{x}\in\Omega, \forall n, n' \in [N], \ \tfrac{1}{\xi} \leq \tfrac{Q_n(\boldsymbol{x})}{Q_{n'}(\boldsymbol{x})} \leq \xi. \tag{5}$$

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

2.2. General Optimization & Learning Framework

Given $Q(\boldsymbol{x};\boldsymbol{w})$, to simulate $\boldsymbol{x}$, the one-run variance for $\mathcal{E}_n$ is:

$$\sigma_n^2(\boldsymbol{w}) \triangleq \mathbb{V}_{\boldsymbol{x}\sim Q}\left[\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\tfrac{P(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}\right]. \tag{6}$$

We evaluate the efficiency of the overall simulation associated with the mixture parameter $\boldsymbol{w}$ by the cost measure $L(\boldsymbol{\sigma}(\boldsymbol{w}))\in\mathbb{R}$ (please refer to Section 2.3 for some examples), where $\boldsymbol{\sigma}(\boldsymbol{w}) \triangleq (\sigma_1(\boldsymbol{w}), \sigma_2(\boldsymbol{w}), ..., \sigma_N(\boldsymbol{w}))$. We formulate the

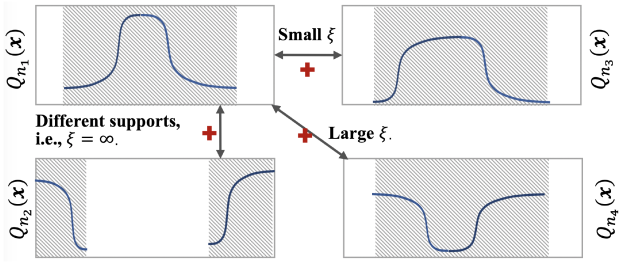


Figure 1: Examples of different level ξ-similarities: an infinite ξ for $\{Q_{n_1}(\boldsymbol{x}), Q_{n_2}(\boldsymbol{x})\}$ implies that even the optimal mixture $Q(\boldsymbol{x};\boldsymbol{w}^*)$ would not work for both $\mathcal{E}_{n_1}$ and $\mathcal{E}_{n_2}$; a large ξ for $\{Q_{n_1}(\boldsymbol{x}), Q_{n_4}(\boldsymbol{x})\}$ implies a slow convergence to $Q(\boldsymbol{x};\boldsymbol{w}^*)$; a small ξ for $\{Q_{n_1}(\boldsymbol{x}), Q_{n_3}(\boldsymbol{x})\}$ implies a fast convergence to $Q(\boldsymbol{x};\boldsymbol{w}^*)$.

following importance sampling optimization problem.

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

$$\min_{\boldsymbol{w}\in\Delta} L(\boldsymbol{\sigma}(\boldsymbol{w})). \tag{7}$$

In general, Problem 1 is a non-linear optimization problem. One challenge in solving Problem 1 is that both $L(\boldsymbol{\sigma}(\boldsymbol{w}))$ and $\nabla L(\boldsymbol{\sigma}(\boldsymbol{w}))$ are computational expensive to compute, i.e., the exact computational complexities are $O(2^M)$ due to the large state space of $\boldsymbol{x}$. To overcome this challenge, we formulate the following framework to estimate (or to perform online learning) the optimal mixture $\boldsymbol{w}$ from samples.

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

- *Select an arm (or event) $I_t\in[N]$ based on an algorithm $\mathcal{A}$ and the sample history $\{(I_0,\boldsymbol{x}^{(0)}),\ldots,(I_{t-1},\boldsymbol{x}^{(t-1)})\}$;*
- *Draw an occurrence profile $\boldsymbol{x}^{(t)}$ from $Q_{I_t}(\boldsymbol{x})$;*
- *Update the proportions of selecting arms (or events) which denoted by $\boldsymbol{w}^{(t)}=(w_1^{(t)},...,w_N^{(t)})$, where $\boldsymbol{w}^{(t)}=\frac{1}{t}\sum_{s\in[t]}\boldsymbol{e}_{I_s}$;*

Objective: Design an algorithm $\mathcal{A}$ to achieve a low and sublinear regret

$$R_T \triangleq L(\boldsymbol{\sigma}(\boldsymbol{w}^{(T)})) - \min_{\boldsymbol{w}\in\Delta} L(\boldsymbol{\sigma}(\boldsymbol{w})). \tag{8}$$

In Problem 2, each arm (or event) x corresponds a pure importance distribution $Q_n(\boldsymbol{x})$. Problem 2 considers general cost function $L(\boldsymbol{\sigma}(\boldsymbol{w}^{(T)}))$. In the following we consider two important instances of $L(\boldsymbol{\sigma}(\boldsymbol{w}^{(T)}))$.

2.3. Two Measures of the MIS Learning Problem

Given $Q(\boldsymbol{x};\boldsymbol{w})$ to simulate $\boldsymbol{x}$, let $\ell_n(\boldsymbol{w})$ 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 $\ell_{max}(\boldsymbol{w})$ measure the simulation cost to achieve desired estimation accuracies for all $\{\mu_n\}_{n=1}^N$. Then:

$$\ell_n(\boldsymbol{w}) \triangleq \frac{\sigma_n^2(\boldsymbol{w})}{\delta_n^2} \text{ and } \ell_{max}(\boldsymbol{w}) \triangleq \max_{n\in[N]} \ell_n(\boldsymbol{w}). \tag{9}$$

Next, we consider the cost measure $L(\boldsymbol{\sigma}(\boldsymbol{w}))$ with various accuracy requirements $\{\delta_n\}_{n=1}^N$, and introduce the corresponding MIS-Learning problems.

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

$$\min_{\boldsymbol{w}\in\Delta}\sum_{n\in[N]}\ell_n(\boldsymbol{w}) \iff \min_{\boldsymbol{w}\in\Delta}\sum_{n\in[N]}\sigma_n^2(\boldsymbol{w})$$
$$\iff \min_{\boldsymbol{w}\in\Delta}\sum_{n\in[N]}\sigma_n^2(\boldsymbol{w})+\mu_n^2.$$

We can define the total loss (cost measure) in terms of the sum of one-run variances as follows:

$$L(\boldsymbol{\sigma}(\boldsymbol{w}))=\sum_{n\in[N]}\sigma_n^2(\boldsymbol{w})+\mu_n^2 \triangleq L_{\text{SumVar}}(\boldsymbol{w}), \tag{10}$$

and name the MIS-Learning with cost measure in Eq. (10) as minimizing the sum of variances (SumVar) MIS-Learning.

MIS-Learning to Minimize Simulation Cost: Consider $\{\mathcal{E}_n\}_{n=1}^N$ with heterogenous accuracy requirements. Specifically, 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's requirement and we want to check whether the undelivery probability is below some threshold o_n. *To accurately state whether $\mu_n\le o_n$ or not*, the CI width should not exceed $\delta_n=|\mu_n-o_n|$. Then:

$$\min_{\boldsymbol{w}\in\Delta}\ell_{max}(\boldsymbol{w}) \iff \min_{\boldsymbol{w}\in\Delta}\max_{n\in[N]}\frac{\sigma_n^2(\boldsymbol{w})}{(\mu_n-o_n)^2}. \tag{11}$$

We define the total loss in terms of the simulation cost to achieve all the desired estimation accuracies as follows:

$$L(\boldsymbol{\sigma}(\boldsymbol{w}))=\max_{n\in[N]}\frac{\sigma_n^2(\boldsymbol{w})}{(\mu_n-o_n)^2} \triangleq L_{\text{SimCos}}(\boldsymbol{w}), \tag{12}$$

and name the MIS-Learning with cost measure in Eq. (12) as minimizing the simulation cost (MinCos) MIS-Learning.

3. Learning to Minimize the Sum of Variances

In this section, we design an algorithm to learn the optimal mixture to minimize the sum of variances. We also present the regret upper bound of our algorithm, which quantifies the impact of the ξ-similarity on the learning speed, and we present the key idea of our proof.

3.1. The Learning Algorithm Design

The idea of our learning algorithm is that at each round: (1) *we first estimate the gradient $\nabla L_{\text{SumVar}}(\boldsymbol{w})$ via historical data samples*; (2) *then we select the arm (or event) based on the estimated gradient*.

Gradient estimation. Consider round t, we aim to estimate $\nabla L_{\text{SumVar}}(\boldsymbol{w}^{(t-1)})$ from historical data samples. Let us first derive $\nabla L_{\text{SumVar}}(\boldsymbol{w})$ as:

$$\nabla L_{\text{SumVar}}(\boldsymbol{w})=\nabla\Big\{\sum_{n\in[N]}\mathbb{E}_{\boldsymbol{x}\sim Q(\boldsymbol{x};\boldsymbol{w})}\Big[\frac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q^2(\boldsymbol{x};\boldsymbol{w})}\Big]\Big\}$$
$$=-\sum_{n\in[N]}\sum_{\boldsymbol{x}\in\Omega}\Big[\frac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q^2(\boldsymbol{x};\boldsymbol{w})}\Big](Q_1(\boldsymbol{x}),...,Q_N(\boldsymbol{x}))$$
$$=\mathbb{E}_{\boldsymbol{x}\sim Q(\boldsymbol{x};\boldsymbol{w})}[(-Z_1(\boldsymbol{x}),\ldots,-Z_N(\boldsymbol{x}))], \tag{13}$$

Algorithm 1 SumVar MIS-Learning

Input: $N, \boldsymbol{w}{=}(\frac{1}{N}, ..., \frac{1}{N}), c_n^{(t)}, \forall n{\in}[N], t{=}1, ..., T$
for all $t{\le}N$ **do**
  Draw $\boldsymbol{x}^{(t)}$ according to distribution $Q_t(\boldsymbol{x})$.
  Record history: $Q_t(\boldsymbol{x}^{(t)})$ and $\mathbf{1}_{\mathcal{E}_t}(\boldsymbol{x}^{(t)})$.
end for
for all $t{>}N$ **do**
  Estimate the gradient $\nabla L_{\mathrm{SumVar}}\left(\boldsymbol{w}^{(t-1)}\right)$ using $\boldsymbol{g}^{(t)}$ derived by Eq. (14).
  Compute the LCB $\underline{\boldsymbol{g}}^{(t)}$, where $\underline{g}_n^{(t)}{=}g_n^{(t)}{-}c_n^{(t)}$.
  Select $I_t{\in}\operatorname{argmin}_{n\in[N]} \underline{g}_n^{(t)}$ and draw $\boldsymbol{x}^{(t)}$ from $Q_{I_t}(\boldsymbol{x})$.
  Record history: $Q_{I_t}(\boldsymbol{x}^{(t)})$ and $\mathbf{1}_{\mathcal{E}_{I_t}}(\boldsymbol{x}^{(t)})$.
  Update $\boldsymbol{w}^{(t)}{\leftarrow}\boldsymbol{w}^{(t-1)}{+}\frac{1}{t}(\boldsymbol{e}_{I_t}{-}\boldsymbol{w}^{(t-1)})$.
end for

where the function $Z_n(\cdot)$ is defined as

$$Z_n(\boldsymbol{x}) \triangleq \frac{P^2(\boldsymbol{x})\sum_{i\in[N]}\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})}{Q^3(\boldsymbol{x};\boldsymbol{w}^{(t-1)})}Q_n(\boldsymbol{x}), \qquad \forall n \in [N].$$

If the historical data samples $\{\boldsymbol{x}^{(s)}\}_{s=1}^{t-1}$ were IID samples of $\boldsymbol{x}{\sim}Q(\boldsymbol{x};\boldsymbol{w}^{(t-1)})$, then the gradient $\nabla L_{\mathrm{SumVar}}(\boldsymbol{w}^{(t-1)})$ can be estimated by $\boldsymbol{g}^{(t)}$, where:

$$g_n^{(t)}{=}\frac{-1}{t-1}\sum_{s\in[t-1]} Z_n(\boldsymbol{x}^{(s)}), \qquad \forall n \in [N]. \tag{14}$$

However, the challenge is that $\{\boldsymbol{x}^{(s)}\}_{s=1}^{t-1}$ are generated from $\boldsymbol{x}^{(s)}{\sim}Q_{I_s}(\boldsymbol{x})$. To address this challenge, the following theorem proves that Eq. (14) is asymptoticallly accurate in estimating the gradient $\nabla L_{\mathrm{SumVar}}(\boldsymbol{w}^{(t-1)})$.

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

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

Arm selection. We outline the arm selection in Algorithm 1. From (Berthet & Perchet, 2017), we know that finding the minimizer of the 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 the corner of simplex Δ that $\min_{\boldsymbol{z}\in\Delta} \boldsymbol{z}^\top \underline{g}^{(t)}$, which is precisely the Frank-Wolfe algorithm (Frank & Wolfe, 1956). Hence, we apply the LCB Frank-Wolfe algorithm to select the arm based on the estimated gradient in Eq. (14). Note that in Algorithm 1, one can select $c_n^{(t)}$ to control the exploration and exploitatoin tradeoffs. Selecting the $c_n^{(t)}$ is closely related to the regret of Algorithm 1. We thus delay the selection in the next subsection, where we analyze the regret.

3.2. Regret Analysis

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

Strong convexity and smoothness of $L_{\mathrm{SumVar}}(\boldsymbol{w})$. 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. f is called α-strongly convex if and only if*

$$\forall \boldsymbol{x}{\in}X, \nabla^2 f(\boldsymbol{x}){\succeq}\alpha I, \tag{15}$$

or equivalently,

$$\forall \boldsymbol{x},\boldsymbol{y}{\in}X, f(\boldsymbol{y}){\ge}f(\boldsymbol{x}){+}\nabla^\top f(\boldsymbol{x})(\boldsymbol{y}{-}\boldsymbol{x}){+}\tfrac{\alpha}{2}\|\boldsymbol{y}{-}\boldsymbol{x}\|_2^2. \tag{16}$$

Similarly, f is called β-smooth if and only if

$$\forall \boldsymbol{x}{\in}X, \nabla^2 f(\boldsymbol{x}){\preceq}\beta I, \tag{17}$$

or equivalently,

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

In the following theorem we prove the strong convexity and smoothness of $L_{\mathrm{SumVar}}(\boldsymbol{w})$.

Theorem 2. *If $\{\mathcal{E}_n\}_{n=1}^N$ has a ξ-similarity, $L_{\mathrm{SumVar}}(\boldsymbol{w})$ given by Eq. (10) is α-strongly convex and β-smooth with:*

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

Remark: Theorem 2 quantifies the impact of ξ-similarities on $L_{\mathrm{SumVar}}(\boldsymbol{w})$. In particular, the strongly convex property of $L_{\mathrm{SumVar}}(\boldsymbol{w})$ may vanish and $L_{\mathrm{SumVar}}(\boldsymbol{w})$ may become 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 as well.

Concentration property of $\boldsymbol{g}^{(t)}$. We aim to characterize how well the estimator $g_n^{(t)}$ concentrates around the gradient $\nabla L_{\mathrm{SumVar}}(\boldsymbol{w}^{(t)})\big|_n$. The 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. (14), the historical data samples $\{\boldsymbol{x}^{(s)}\}_{s=1}^{t-1}$ are not IID. The following theorem resolves this challenge by quantifying a tradeoff between $\zeta^{(t)}$ and $\epsilon_n^{(t)}$.

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

$$\epsilon_n^{(t)}{=}\frac{\ln\frac{1}{\zeta^{(t)}}}{3t}Z_n^{max}{+}\sqrt{\frac{1}{9t^2}(\ln\frac{1}{\zeta^{(t)}}Z_n^{max})^2{+}\frac{2}{t}\ln\frac{1}{\zeta^{(t)}}\mathbb{V}Z_n(\boldsymbol{x})}.$$

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

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

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

Remark: Theorem 3 serves as a building block for one to vary the $\zeta^{(t)}$ and $\epsilon^{(t)}$, to attain different confidence and variation tradeoffs. This confidence and variation tradeoff is essential for us 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 $\mathbb{V}Z_n(\boldsymbol{x}){=}O(\xi^5)$, i.e, the CI width of $\boldsymbol{g}^{(t)}$ is proportional to ξ. This reveals the impact of similarity ξ

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 prove the regret upper bound of it. Due to page limit, we present the sketch proof in the next subsection and leave the detailed proof in the appendix.

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 $L_{\text{SumVar}}(\boldsymbol{w})$ in Eq. (10), after T steps of the SumVar algorithm, the choice of $c_n^{(t)}{=}\epsilon_n^{(t)}$ and*

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

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

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

otherwise, we have:

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

where,

$$C_1=O\Big(\tfrac{N^2(\ln T_0)^2\xi^6}{\alpha\eta^2}+\tfrac{N^{3/2}\xi^2\sum_{i\in[N]}\mu_i}{T_0}+\tfrac{N\ln T_0\beta\xi^3}{\alpha\eta^2}\Big),$$

$$C_3=O\Big(\tfrac{N^{3/2}\xi^2\sum_{i\in[N]}\mu_i}{T_0}+\tfrac{N(\ln T_0)^2\beta\xi^2}{\alpha\eta^2}\Big),$$

$$C_2=C_5=O(\beta),\ C_4=O\Big(\tfrac{\sqrt{N\xi^5}\beta}{\alpha\eta^2}\Big),\ C_6=O\Big(\tfrac{N\xi^5}{\alpha\eta^2}\Big).$$

Remark: Theorem 4 shows that the regret bound is proportional to the ξ-similarity. It also reveals that a small ξ implies a fast convergence to the optimal mixture $Q(\boldsymbol{x};\boldsymbol{w})$.

3.3. Proof Sketch

Here we state the sketch proof for Theorem 4, in which we discuss the regret upper bound of the SumVar algorithm.

Let I_t stand for the index of arm selected by $\mathcal{A}$ at round t. Recall that $\boldsymbol{w}^{(t)}$ is defined as the proportions of arm selections, i.e., $\boldsymbol{w}^{(t)}{=}\frac{1}{t}\sum_{s\in[t]}\boldsymbol{e}_{I_s}$. We can derive the following recurrence:

$$\boldsymbol{w}^{(t+1)}=\tfrac{t\boldsymbol{w}^{(t)}+\boldsymbol{e}_{I_{t+1}}}{t+1}=\boldsymbol{w}^{(t)}+\tfrac{\boldsymbol{e}_{I_{t+1}}-\boldsymbol{w}^{(t)}}{t+1}. \tag{24}$$

Let $\boldsymbol{w}^*$ be the optimal weighting strategy:

$$\boldsymbol{w}^*=\text{argmin}_{\boldsymbol{w}\in\Delta}\,L_{\text{SumVar}}(\boldsymbol{w}), \tag{25}$$

and define $\boldsymbol{e}_{*_{t+1}}$ as the following minimizer:

$$\boldsymbol{e}_{*_{t+1}}=\text{argmin}_{\boldsymbol{z}\in\Delta}\,\boldsymbol{z}^\top\nabla L_{\text{SumVar}}(\boldsymbol{w}^{(t)}), \tag{26}$$

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

$$\varepsilon^{(t+1)}=\nabla L_{\text{SumVar}}(\boldsymbol{w}^{(t)})^\top(\boldsymbol{e}_{I_{t+1}}-\boldsymbol{e}_{*_{t+1}}). \tag{27}$$

The proof of Theorem 4 can be broken down into the following steps.

Step 1: By the convexity and smoothness properties of $L_{\text{SumVar}}(\boldsymbol{w})$ in Theorem 2, we first partition the regret R_T and show that

$$R_T=\tfrac{1}{T}\Big[\textstyle\sum_{s\in[T]}\frac{\beta}{s}+\sum_{s\in[T]}\varepsilon^{(s)}\Big]\leq\beta\frac{\ln T}{T}+\frac{\sum_{s\in[T]}\varepsilon^{(s)}}{T}. \tag{28}$$

Step 2: To bound R_T, it is equivalent to bound $\sum_{s\in[T]}\varepsilon^{(s)}$. We start by looking at $c_n^{(t)}$, i.e., the confidence bound when estimating $\nabla L_{\text{SumVar}}\left(\boldsymbol{w}^{(t)}\right)\big|_n$ with $g_n^{(t)}$, which affects the accuracy of estimating $\boldsymbol{e}_{*_t}$ with $\boldsymbol{e}_{I_t}$ when $n{=}I_t$. The following claim reveals the relationship between $c_{I_t}^{(t)}$ and $\varepsilon^{(t+1)}$:

Claim 1. *Assume $c_n^{(t)}$ satisfies*

$$\mathbb{P}\Big[g_n^{(t)}-\nabla L_{\text{SumVar}}\big(\boldsymbol{w}^{(t)}\big)\big|_n\leq c_n^{(t)}\Big]\leq\zeta^{(t)}, \tag{29}$$

$$\mathbb{P}\Big[g_n^{(t)}-\nabla L_{\text{SumVar}}\big(\boldsymbol{w}^{(t)}\big)\big|_n\geq -c_n^{(t)}\Big]\leq\zeta^{(t)}. \tag{30}$$

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

We then derive the expression of $c_n^{(t)}$. By Theorem 3, we know that Equation (29) and (30) can be satisfied with the choice of $c_n^{(t)}{=}\epsilon_n^{(t)}$ where $\epsilon_n^{(t)}$ is defined in Theorem 3.

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

Claim 2. *With the choice of $c_n^{(t)}{=}\epsilon_n^{(t)}$, we have:*

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

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

Combine Claim 1, 2 and Theorem 3, we show that with the choice of $c_n^{(t)}$ in Theorem 3, 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 2.

Step 3: Next we show 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 $O\left(\frac{1}{T}\sum_t c_{I_t}^{(t)}\right)$.

Let η be the distance from $\boldsymbol{w}^*$ to $\partial\Delta$. Let $c^{(t)}\triangleq\max_{n\in[N]}c_n^{(t)}$. Claim 2 provides an upper bound of $c_n^{(t)}$. Next, we utilize $c^{(t)}$ to bound R_T:

Claim 3. *Assume we select $\zeta^{(t)}$ properly such that $\frac{1}{t}\sum_{s\in[t]}\left(c^{(t)}\right)^2\geq\left(c^{(t+1)}\right)^2$. Then with a probability at least $1-N\sum_t\zeta^{(t)}$ that:*

$$TR_T\leq\tfrac{\alpha\eta^2}{2}+\tfrac{\pi^2\beta^2}{3\alpha\eta^2}+\beta\ln T$$
$$+\tfrac{8\beta}{\alpha\eta^2}\textstyle\sum_{t\in[T]}\frac{c^{(t)}}{t}+\frac{8}{\alpha\eta^2}\sum_{t\in[T]}\left(c^{(t)}\right)^2. \tag{31}$$

Step 4: We now discuss how to select $\zeta^{(t)}$ to guarantee that $\frac{1}{t}\sum_{s\in[t]}\left(c^{(t)}\right)^2\geq\left(c^{(t+1)}\right)^2$, and give bounds of $\sum_t\left(c^{(t)}\right)^2$ and $\sum_t\frac{c^{(t)}}{t}$ by the following:

Claim 4. *With the choice of* $\zeta^{(t)}=\begin{cases} T_0^{-2}, & \text{if } t\leq T_0; \\ t^{-2}, & \text{if } t>T_0.\end{cases}$

If $c^{(t)} = \frac{4}{3} N\xi^3 \frac{\ln\frac{1}{\zeta^{(t)}}}{t}$, *we have:*

$$\begin{cases} \sum_{t\in[T]} \left(c^{(t)}\right)^2 \leq \frac{64N^2\xi^6}{9}\left[\frac{\pi^2(\ln T_0)^2}{6}+2\right], \\ \sum_{t\in[T]} \frac{c^{(t)}}{t} \leq \frac{8N\xi^3}{3}\left[\frac{\pi^2 \ln T_0}{6}+1\right]. \end{cases} \tag{32}$$

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

$$\begin{cases} \sum_{t\in[T]} (c^{(t)})^2 \leq 4N\xi^5 \sum_m \mu_m \left[(\ln T_0)^2+(\ln T)^2\right], \\ \sum_{t\in[T]} \frac{c^{(t)}}{t} \leq \sqrt{8N\xi^5 \sum_m \mu_m}\Big\{(2+\sqrt{2})\sqrt{\ln T_0} \\ \qquad +\sqrt{2\pi}\,\mathrm{erf}\left(\sqrt{\frac{\ln T}{2}}\right)\Big\}. \end{cases} \tag{33}$$

Step 5: With a probability no more than $N\sum_t \zeta^{(t)} \leq \frac{2N}{T_0}$, we have $R_T \leq \xi^2 \sum_{i\in[N]} \mu_i \sqrt{N}$. With a probability at least $1-N\sum_t \zeta^{(t)}$, we have the bound of R_T in Claim 3. By plugging bounds of $\sum_t (c^{(t)})^2$ and $\sum_t \frac{c^{(t)}}{t}$ into Claim 3, we finish the proof of Theorem 4. ■

4. Learning to Minimize the Simulation Cost

In this section, we design an algorithm to learn the optimal mixture to minimize the simulation cost. We also present the regret upper bound of our algorithm, which quantifies the impact of ξ-similarity on the learning speed.

4.1. The Learning Algorithm Design

We first develop a linear approximation framework to locate a search direction. Then, we design an estimator to estimate the search direction from data samples. Finally, we use the estimated search direction to select the arm.

Search direction. One challenge in locating the search direction is that the objective function $L_{\mathrm{SimCos}}(\boldsymbol{w})$ takes the pointwise maximum of functions $\ell_n(\boldsymbol{w})$, leading to the non-smoothness. Furthermore, one constraint is that Problem 2 implies a step size in updating $\boldsymbol{w}^{(t)}$ is $1/t$, i.e.,

$$\boldsymbol{w}^{(t+1)} = \frac{t\boldsymbol{w}^{(t)}+\boldsymbol{z}}{t+1}, \tag{34}$$

where $\boldsymbol{z}\in\Delta$. Namely, to determine the search direction, we need to determine $\boldsymbol{z}$. To measure the potential of $\boldsymbol{z}$ in decreasing $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)})$, we consider a *linearization* of $L_{\mathrm{SimCos}}(\boldsymbol{w})$ at $\boldsymbol{w}=\boldsymbol{w}^{(t)}$:

$$\begin{aligned} L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}) &= \max_{n\in[N]} \ell_n(\boldsymbol{w}^{(t)}) + \nabla\ell_n(\boldsymbol{w}^{(t)})^\top(\boldsymbol{w}^{(t+1)}-\boldsymbol{w}^{(t)}) \\ &= \max_{n\in[N]} \ell_n(\boldsymbol{w}^{(t)}) + \nabla\ell_n(\boldsymbol{w}^{(t)})^\top \frac{\boldsymbol{z}-\boldsymbol{w}^{(t)}}{t+1}, \end{aligned} \tag{35}$$

and bound its approximation error in the following lemma:

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

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 $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t+1)})$. Hence, given $\boldsymbol{w}^{(t)}$, we consider the *minimizer of* $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z})$ as the search direction. Furthermore, the minimum of $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z})$ can be attained by *the standard direction with steepest decrease*, i.e.,

$$\min_{\boldsymbol{z}\in\Delta} L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}) = \min_{\boldsymbol{z}\in\mathcal{U}} L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}),$$

where $\mathcal{U} \triangleq \{\boldsymbol{e}_1, ..., \boldsymbol{e}_N\}$ represents the standard basis. This implies that we can reduce the search space from Δ to $\mathcal{U}$ and simplify the estimation of the search direction as we proceed to show. We take such steepest decrease direction as the search direction, and denote it by:

$$\boldsymbol{e}_t^* = \mathrm{argmin}_{\boldsymbol{z}\in\mathcal{U}} L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}). \tag{36}$$

Search direction estimation. We consider the following equivalent form for the search direction:

$$\boldsymbol{e}_t^* = \mathrm{argmin}_{\boldsymbol{z}\in\mathcal{U}} L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}) - L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)}). \tag{37}$$

The form of search direction is useful to estimate the search direction, for the value of $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}) - L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)})$ 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 $\{L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n)\}_{n=1}^N$ and $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)})$ to locate $\boldsymbol{e}_t^*$. Essentially, we need to estimate $\ell_n(\boldsymbol{w}^{(t)})$ and $\nabla\ell_n(\boldsymbol{w}^{(t)})$ from the data samples $\{\boldsymbol{x}^{(s)}\}_{s=1}^{t-1}$. We have similar challenges as Section 3.1, i.e., data samples are not IID, and we address them with a method similar as in Section 3.1. We estimate $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n) - L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)})$ as $g_n^{(t)}$ where $g_n^{(t)}$ is derived as $g_n^{(t)} \triangleq \dot{g}_n^{(t)} - \ddot{g}^{(t)}$ and:

$$\begin{aligned} \dot{g}_n^{(t)} &= \max_{i\in[N]} \frac{t+1}{t}\widehat{A}_i\left(\boldsymbol{w}^{(t-1)}\right) - \frac{1}{t}\widehat{B}_i\left(\boldsymbol{w}^{(t-1)};n\right) - \frac{(\widehat{\mu}_i^{(t-1)})^2}{(\widehat{\mu}_i^{(t-1)}-o_i)^2}, \\ \ddot{g}^{(t)} &= \max_{i\in[N]} \widehat{A}_i\left(\boldsymbol{w}^{(t-1)}\right) - \frac{(\widehat{\mu}_i^{(t-1)})^2}{(\widehat{\mu}_i^{(t-1)}-o_i)^2}, \\ \widehat{B}_i\left(\boldsymbol{w}^{(t)};n\right) &= \frac{1}{(\widehat{\mu}_i^{(t-1)}-o_i)^2}\frac{1}{t}\sum_{s\in[t]} \frac{P^2\left(\boldsymbol{x}^{(s)}\right)\mathbf{1}_{\mathcal{E}_i}\left(\boldsymbol{x}^{(s)}\right)}{Q^2\left(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)}\right)}\frac{Q_n\left(\boldsymbol{x}^{(s)}\right)}{Q\left(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)}\right)}, \\ \widehat{A}_i\left(\boldsymbol{w}^{(t)}\right) &= \frac{1}{(\widehat{\mu}_i^{(t-1)}-o_i)^2}\cdot\frac{1}{t}\sum_{s\in[t]} \frac{P^2(\boldsymbol{x}^{(s)})\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x}^{(s)})}{Q^2\left(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)}\right)}; \end{aligned} \tag{38}$$

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

Theorem 5. *Consider the MIS-Learning framework, where at round* $t, t\in[T]$ *take the* I_t*-th distribution* $Q_{I_t}(\boldsymbol{x})$ *to generate* $\boldsymbol{x}^{(t)}$. *Then, both* $\lim_{t\to\infty}\|\dot{g}_n^{(t)} - L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n)\|=0$ *and* $\lim_{t\to\infty}\|\ddot{g}^{(t)} - L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)})\|=0$.

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

Arm selection. Based on $g_n^{(t)}, n\in[N]$, we estimate the steepest search direction using the LCB framework and we outline it in Algorithm 2. Selecting the $c_n^{(t)}$ is closely related to the regret of Algorithm 2. We will discuss the selection in the next subsection, where we analyze the regret.

4.2. Regret Analysis

To first decompose the regret. Denote the optimal mixture as $\boldsymbol{w}^*$, the optimal search direction as $\boldsymbol{e}_t^*$, and the estimated search direction (i.e., the action direction) as $\boldsymbol{e}_{I_t}$. Then we

Algorithm 2 SimCos MIS-Learning

Input: N, $\boldsymbol{w}=(\frac{1}{N},...,\frac{1}{N})$
for all $t\leq N$ **do**
  Draw $\boldsymbol{x}^{(t)}$ according to distribution $Q_t(\boldsymbol{x})$.
  Record history: $Q_n(\boldsymbol{x}^{(t)})$ and $\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x}^{(t)}), n\in[N]$.
end for
for all $t>N$ **do**
  Estimate $\mu_n^{(t-1)}, n\in[N]$ by

$$\widehat{\mu}_n^{(t-1)}=\frac{1}{t-1}\sum_{s\in[t-1]}\frac{P\big(\boldsymbol{x}^{(s)}\big)\mathbf{1}_{\mathcal{E}_n}\big(\boldsymbol{x}^{(s)}\big)}{Q\big(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(s)}\big)}. \quad (39)$$

  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. (38).
  Compute the LCB $\underline{g}_n^{(t)}$, where $\underline{g}_n^{(t)}=g_n^{(t)}-c_n^{(t)}$.
  Select arm $I_t\in\operatorname{argmin}_{n\in[N]}\underline{g}_n^{(t)}$.
  Update $\boldsymbol{w}^{(t)}\leftarrow\boldsymbol{w}^{(t-1)}+\frac{1}{t}(e_{I_t}-\boldsymbol{w}^{(t-1)})$.
end for

decompose the regret as:

$$\begin{aligned}
&L_{\rm SimCos}(\boldsymbol{w}^{(t+1)})-L_{\rm SimCos}(\boldsymbol{w}^*)\\
&\leq L_{\rm SimCos}\big(\tfrac{t\boldsymbol{w}^{(t)}+\boldsymbol{e}_{I_t}}{t+1}\big)-L_{\rm SimCos}\big(\tfrac{t\boldsymbol{w}^{(t)}+\boldsymbol{e}_t^*}{t+1}\big) \quad \text{(R1)}\\
&+L_{\rm SimCos}\big(\tfrac{t\boldsymbol{w}^{(t)}+\boldsymbol{e}_t^*}{t+1}\big)-L_{\rm SimCos}\big(\tfrac{t\boldsymbol{w}^{(t)}+\boldsymbol{w}^*}{t+1}\big) \quad \text{(R2)}\\
&+L_{\rm SimCos}\big(\tfrac{t\boldsymbol{w}^{(t)}+\boldsymbol{w}^*}{t+1}\big)-L_{\rm SimCos}(\boldsymbol{w}^*). \quad \text{(R3)}
\end{aligned}$$

This decomposition has three parts. Part $R1$ is the *estimation error*, which is essentially governed by the concentration of $g_n^{(t)}$ in estimating $L_{\rm SimCos}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n)-L_{\rm SimCos}(\boldsymbol{w}^{(t)})$. Part $R2+R3$ is the *approximation error*, which is essentially governed by the convexity and smoothness of the objective function $L_{\rm SimCos}(\boldsymbol{w})$.

In the remainder of this section, similar as in the SumVar case, we first establish two building blocks: (1) *The strong convexity and smoothness of $L_{\rm SimCos}(\boldsymbol{w})$ and its components*; (2) *The concentration property of $\boldsymbol{g}^{(t)}$ in estimating* $L_{\rm SimCos}(\boldsymbol{w}^{(t)};\boldsymbol{z})-L_{\rm SimCos}(\boldsymbol{w}^{(t)})$. Then we apply these two blocks to derive the regret upper bound of Algorithm 1.

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

Corollary 1. *If $\{\mathcal{E}_n\}_{n=1}^N$ has a ξ-similarity, $\ell_n(\boldsymbol{w}), n\in[N]$ in Eq.(9) is α_n-strongly convex and β_n-smooth, where*

$$\alpha_n=\frac{2\mu_n^2}{\xi^2(\mu_n-o_n)^2} \text{ and } \beta_n=\frac{2\xi^3\mu_n}{(\mu_n-o_n)^2}. \quad (40)$$

Such convexity and smoothness of $\ell_n(\boldsymbol{w}), n\in[N]$ guarantee the convexity of $L_{\rm SimCos}(\boldsymbol{w})$:

Corollary 2. *If $\{\mathcal{E}_n\}_{n=1}^N$ has a ξ-similarity, $L_{\rm SimCos}(\boldsymbol{w})$ in Eq. (12) is α'-strongly convex, where*

$$\alpha'\triangleq\min_{n\in[N]}\alpha_n=\min_{n\in[N]}\frac{2\mu_n^2}{\xi^2(\mu_n-o_n)^2}. \quad (41)$$

Remark: Corollary 1 and 2 quantify the impact of the ξ-similarity on the convexity and smoothness of $L_{\rm SumVar}(\boldsymbol{w})$ and its components. We also need to point out that the tight approximation mentioned in Lemma 1 is also guaranteed by the convexity and smoothness of $\ell_n(\boldsymbol{w}), n\in[N]$.

Concentration property of $g_n^{(t)}$. In the following theorem, we characterize how well the estimator $g_n^{(t)}$ concentrates around the $L_{\rm SimCos}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n)-L_{\rm SimCos}(\boldsymbol{w}^{(t)})$.

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

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

where $A_i(\boldsymbol{x};\boldsymbol{w}^{(t)})=\frac{1}{(\mu_i-o_i)^2}\frac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})}{Q^2\big(\boldsymbol{x};\boldsymbol{w}^{(t)}\big)}$,

$$B_i(\boldsymbol{x};\boldsymbol{w}^{(t)};n)=\frac{1}{(\mu_i-o_i)^2}\frac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})Q_n(\boldsymbol{x})}{Q^3\big(\boldsymbol{x};\boldsymbol{w}^{(t)}\big)},$$

$$a_i^{(t)}=\varphi\left(A_i\big(\boldsymbol{x};\boldsymbol{w}^{(t)}\big)\right), b_{i,n}^{(t)}=\varphi\left(B_i\big(\boldsymbol{x};\boldsymbol{w}^{(t)};n\big)\right),$$

$$C_1=\max_{k\in[N]}\frac{2\xi^2\mu_k}{(\mu_k-o_k)^2}+\max_{k\in[N]}\frac{2\xi^3}{(\mu_k-o_k)^2}.$$

Then, it holds that

$$\mathbb{P}\Big[g_n^{(t)}-(L_{\rm SimCos}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n)-L_{\rm SimCos}(\boldsymbol{w}^{(t)}))\leq\epsilon_n^{(t)}\Big]\leq\zeta^{(t)},$$
$$\mathbb{P}\Big[g_n^{(t)}-(L_{\rm SimCos}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n)-L_{\rm SimCos}(\boldsymbol{w}^{(t)}))\geq-\epsilon_n^{(t)}\Big]\leq\zeta^{(t)}.$$

Remark: We need to point out that $C_1=O(\xi^3)$ and $a_i^{(t)}+b_{i,n}^{(t)}$ $=O(\sqrt{\xi^3\frac{\ln(8/\zeta^{(t)})}{t}})$. Hence, 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 for Algorithm 2 and prove its regret upper bound. Due to page limit, we present the sketch proof in the next subsection and leave the detailed proof in the appendix.

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 $L_{\rm SimCos}$ in Eq. (12), after T steps of the SimCos algorithm, the choice of $c_n^{(t)}=\epsilon_n^{(t)}$ and*

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

holds the following:

$$\begin{aligned}\mathbb{E}R_T\leq\ &O(\xi^3)\tfrac{1}{T}+O(\beta'+\xi^3)\tfrac{\ln T}{T}\\ &+O(\xi^3)\tfrac{(\ln T)^2}{T}+O(\xi^{5/2})\sqrt{\tfrac{\ln T}{T}}.\end{aligned} \quad (42)$$

Remark: Theorem 7 reveals that: The regret bound of SimCos algorithm is proportional to the ξ-similarity, and; A small ξ implies a fast convergence to the optimal mixture.

4.3. Proof Sketch

Here we state the sketch proof for Theorem 7, in which we discuss the regret upper bound of the SimCos algorithm..

Still, let $\boldsymbol{w}^*$ as the optimal mixture:

$$\boldsymbol{w}^* = \operatorname{argmin}_{\boldsymbol{w}\in\Delta} L_{\text{SimCos}}(\boldsymbol{w}). \tag{43}$$

Recall that $\boldsymbol{e}_{*_t}$, i.e., the steepest decent direction of the linearization $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z})$, is reorganized in Equation (**??**). We estimate $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)})$ as $g_n^{(t)}$ and so estimate $\boldsymbol{e}_{*_t}$ as $\boldsymbol{e}_{I_t}$ where $I_t = \operatorname{argmin}_{n\in[N]} g_n^{(t)}$. The proof of Theorem 4 can be broken down into the following steps.

Step 1: We first partition the regret as:

$$L_{\text{SimCos}}(\boldsymbol{w}^{(t+1)}) - L_{\text{SimCos}}(\boldsymbol{w}^*) \leq R1 + R2 + R3, \tag{44}$$

where part $R1$, $R2$ and $R3$ are given in Section 4.2. In the following, we will bound each part of the regret.

Step 2: By the definition of linearization $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z})$, and the strong convexity and smoothness properties of $\ell_n(\boldsymbol{w})$ given in Corollary 1, we have:

$$R1 \leq L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{I_t}) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t}) + \frac{\beta'}{2}\left\|\frac{\boldsymbol{e}_{I_t} - \boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 - \frac{\alpha'}{2}\left\|\frac{\boldsymbol{e}_{*_t} - \boldsymbol{w}^{(t)}}{t+1}\right\|_2^2, \tag{45}$$

where $\alpha' = \min_{n\in[N]}\alpha_n$, $\beta' = \max_{n\in[N]}\beta_n$. To bound $R1$, we then look at $c_n^{(t)}$, i.e., the confidence bound when estimating $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)})$ with $g_n^{(t)}$, which affects the accuracy of estimating $\boldsymbol{e}_{*_t}$ with $\boldsymbol{e}_{I_t}$ when $n = I_t$. The relationship between $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{I_t}) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t})$ and $c_{I_t}^{(t)}$ can be revealed by the following claim:

Claim 5. *Assume $c_n^{(t)}$ satisfies*

$$\mathbb{P}\Big[g_n^{(t)} - \Big(L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{I_t}) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t})\Big) \leq c_n^{(t)}\Big] \leq \zeta^{(t)}, \tag{46}$$

$$\mathbb{P}\Big[g_n^{(t)} - \Big(L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{I_t}) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t})\Big) \geq -c_n^{(t)}\Big] \leq \zeta^{(t)}. \tag{47}$$

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

$$L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{I_t}) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t}) \leq 2c_{I_{t+1}}^{(t)}.$$

By the above discussion, we bound R_1 by the following:

$$R1 \leq 2c_{I_t}^{(t)} + \frac{\beta'}{2}\left\|\frac{\boldsymbol{e}_{I_t} - \boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 - \frac{\alpha'}{2}\left\|\frac{\boldsymbol{e}_{*_t} - \boldsymbol{w}^{(t)}}{t+1}\right\|_2^2. \tag{48}$$

Step 3: By the definition of $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z})$ and the optimality of $\boldsymbol{e}_{*_t}$, we show that $R2$ is upper bounded by:

$$R2 \leq -\frac{\alpha'}{2}\left\|\frac{\boldsymbol{w}^* - \boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 + \frac{\beta'}{2}\left\|\frac{\boldsymbol{e}_{*_t} - \boldsymbol{w}^{(t)}}{t+1}\right\|_2^2. \tag{49}$$

Step 4: By the strong convexity and smoothness properties of $\ell_n(\boldsymbol{w})$, we show that $R3$ is upper bounded by:

$$R3 \leq \frac{t}{t+1}\left[L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)}\right) - L_{\text{SimCos}}(\boldsymbol{w}^*)\right] + \frac{\beta'-\alpha'}{2}\left\|\frac{\boldsymbol{w}^* - \boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 \tag{50}$$

Step 5: Combine the upper bounds of part R_1, R_2 and $R3$, we show that:

$$R_T \leq \frac{2}{T}\sum_{t\in[T-1]} c_{I_t}^{(t)} + 3(\beta'-\alpha')\frac{\ln\frac{T}{2}}{T}. \tag{51}$$

Step 6: Next, we focus on bounding $c_n^{(t)}$, which measures the accuracy in estimating $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)})$ with $g_n^{(t)}$. By Theorem 6, we have the upper bounds of $c_n^{(t)}$ under different conditions which depend on C_1, $a_i^{(t)}$ and $b_{i,n}^{(t)}$. The bound of C_1 is given by Theorem 6. Next, we analyze the bounds of $a_i^{(t)}$ and $b_{i,n}^{(t)}$:

$$a_i^{(t)} = \varphi\left(A_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right)\right) \leq \frac{2\xi^2}{3(\mu_i - o_i)^2}\frac{\ln\frac{8}{\zeta^{(t)}}}{t} + \frac{\sqrt{2\xi^3\mu_i}}{(\mu_i - o_i)^2}\sqrt{\frac{\ln\frac{8}{\zeta^{(t)}}}{t}}; \tag{52}$$

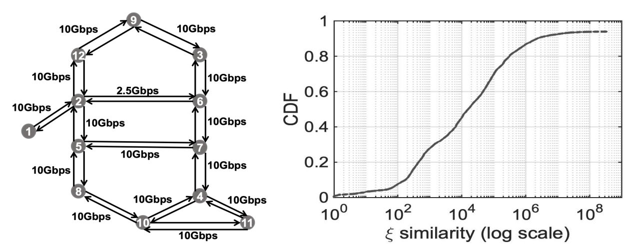
$$b_{i,n}^{(t)} = \varphi\left(B_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)};n\right)\right) \leq \frac{2\xi^3}{3(\mu_i - o_i)^2}\frac{\ln\frac{8}{\zeta^{(t)}}}{t} + \frac{\sqrt{2\xi^5\mu_i}}{(\mu_i - o_i)^2}\sqrt{\frac{\ln\frac{8}{\zeta^{(t)}}}{t}}. \tag{53}$$

Combining Theorem 6, and Equation (51), (52) and (53), we can finish the proof of Theorem 7. ∎

5. Experiment results

We evaluate our methods on the Abilene network (Abi; Jiang et al., 2009), and we aim at efficiently evaluating the impact of *network link failures* on the occurrences of interested events $\mathcal{E}_n, n\in[N]$, which are specified as the *non-satisfaction of bandwidth demands for traffic flows n,* $n\in[N]$.

As depicted in Fig. 2(a), the network contains 12 nodes and 30 links. The occurrences of link failures are indicated by $\boldsymbol{x}$, and each link fails with a probability of 0.01. The topology and traffic matrices are collected from (Zhang, 2014). There are 132 competing flows, and their bandwidth demands are extracted from (Liu & Lui, 2019). 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 (Jain et al., 2013).



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

Figure 2: The network topology and ξ-similarity information.

For each interested event $\mathcal{E}_n$,we take the *customized pure IS distribution proposed in (Liu & Lui, 2019)* as the efficient IS distribution $Q_n(\boldsymbol{x})$ for $\mathcal{E}_n$. To accurately estimate $\{\mu_n\}$ for a group of events $\{\mathcal{E}_n\}$, authors in (Liu & Lui, 2019) considers the MIS solution with a *uniform mixture* strategy $\boldsymbol{w} = (\frac{1}{N}, ..., \frac{1}{N})$. In the following, we apply our MIS-Learning framework to learn a more efficient mixture strategy $\boldsymbol{w}^*$ to minimize the cost measure $L(\boldsymbol{\sigma}(\boldsymbol{w}))$.

We first derive the ξ-similarity 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. 2(b). By setting the upper thresholds of the pairwise ξ-similarity, we can partition $\{\mathcal{E}_n\}_{n=1}^{N}$ into different subsets, on which we will apply our MIS-Learning method to find a mixture strategy $\boldsymbol{w}$ to efficiently estimate their occurrence probabilities simultaneously. By setting the upper bounds of the pointwise ξ-similarity as $\xi \leq 100$, $\xi \leq 200$, $\xi \leq 300$ and $\xi \in [1000, 5000]$, we obtain interested event sets $\{\mathcal{E}_n\}_{n=1}^{N'}$ with set sizes $N'=16$, $N'=19$, $N'=30$ and $N'=5$.

Minimizing the sum of variances.We start with the SumVar MIS-Learning with $L(\boldsymbol{\sigma}(\boldsymbol{w})) \triangleq L_{\text{SumVar}}(\boldsymbol{w})$. For each interested event set $\{\mathcal{E}_n\}_{n=1}^{N'}$ with the corresponding ξ-similarity threshold, we run the SumVar MIS-Learning for 80,000 runs and plot the cost measure $L_{\text{SumVar}}(\boldsymbol{w})$ of each round in Fig. 3, and compare with the uniform mixture strategy proposed in (Liu & Lui, 2019). Fig. 3(a), 3(b) and 3(c) show the reduction of $L_{\text{SumVar}}(\boldsymbol{w})$ achieved by the SumVar MIS-Learning with a small ξ-similarity, and 3(d) shows the performance of 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% when $\xi \leq 100$, $\xi \leq 200$, $\xi \leq 300$ and $\xi \in [1000, 5000]$.

Minimizing the simulation cost. We consider the SimCos MIS-Learning with $L(\boldsymbol{\sigma}(\boldsymbol{w})) \triangleq L_{\text{SimCos}}(\boldsymbol{w})$. For each interested event set $\{\mathcal{E}_n\}_{n=1}^{N'}$ with the corresponding ξ-similarity threshold, we run the SimCos MIS-Learning for 80,000 runs and plot the cost measure $L_{\text{SimCos}}(\boldsymbol{w})$ of each round in Fig. 3, and compare with the uniform mixture. Fig. 3(e), 3(f) and 3(g) show the reduction of $L_{\text{SimCos}}(\boldsymbol{w})$ achieved by the SimCos MIS-Learning with a small ξ-similarity, while 3(h) shows the performance of SimCos MIS-Learning with a large ξ-similarity. The SimCos MIS-Learning with Algorithm 2 reduces the cost measure by 35.7%, 55.1%, 39.9% and 61.6% when $\xi \leq 100$, $\xi \leq 200$, $\xi \leq 300$ and $\xi \in [1000, 5000]$.

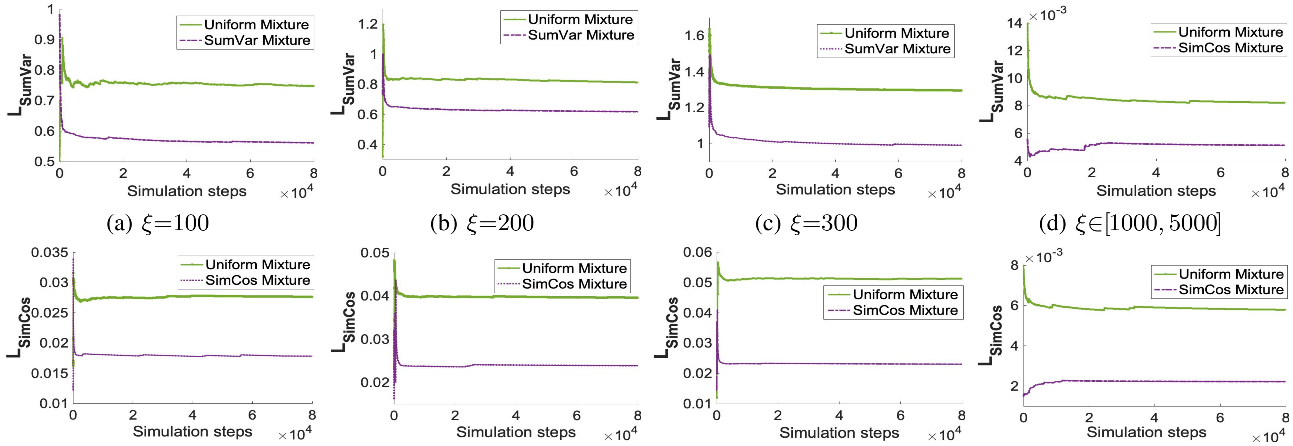
Impact of ξ-similarity on the convergence rate. We take a detailed look at the convergence of cost measure in Fig.4, and compare the convergence rate of the large ξ case (i.e., $\xi \in [1000, 5000]$) and of the small ξ case (i.e., $\xi \leq 300$). For SumVar MIS-Learning with Algorithm 1, Theorem 4 states that the regret $L_{\text{SumVar}}(\boldsymbol{w}) - L_{\text{SumVar}}(\boldsymbol{w}^*)$ first decreases at a fast rate in Eq. (B.56) and then at a slow rate in Eq. (B.57). Theo-rem 4 also reveals that a smaller ξ implies a longer fast rate period. As shown in Fig. 4, with a small ξ, $L_{\text{SumVar}}(\boldsymbol{w})$ 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 $O(\frac{1}{T})$ and then at a slow rate of $Q(\sqrt{\ln T/T})$ in Eq. (C.73). As shown in Fig. 4, with a small ξ, $L_{\text{SimCos}}(\boldsymbol{w})$ decreases first at a fast rate and then at a slow rate; with a large ξ, the short fast rate period vanishes.

6. Related work

MIS vs. IS for rare event simulation: Comprehensive reviews on rare event simulation are given in (Bucklew, 2013; Rubino & Tuffin, 2009). These works are mainly IS based: they estimate the probability of rare event $\mathcal{E}_n$ by simulating the system under an alternative probability measure $Q_n(\boldsymbol{x})$ and then unbias the results (Frank et al., 2008). Given many rare events to estimate, as each $Q_n(\boldsymbol{x})$ 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 the corresponding $Q_n(\boldsymbol{x})$. Authors in (Liu & Lui, 2019) propose a sequential IS method to obtain efficient IS distribution for each single event and then take a uniform mixture to estimate multiple events. In this work, we aim to learn an efficient MIS distribution that works for many rare events. 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 sets with small ξ, which can be efficiently estimated via MIS at the same time.

MIS-Learning vs. Stochastic Optimization: The MIS-Learning can be viewed as the stochastic optimization over the simplex: to minimize the objective $L(\boldsymbol{\sigma}(\boldsymbol{w}))$, we choose at each round an action I_t, which affects the variable $\boldsymbol{w}$ and provides us observations on $L(\boldsymbol{\sigma}(\boldsymbol{w}))$. In the common case where objectives are *smooth*, i.e., $L(\boldsymbol{\sigma}(\boldsymbol{w})) \triangleq L_{\text{SumVar}}(\boldsymbol{w})$, iterative gradient-based methods, such as gradient descent and stochastic gradient descent (Papamakarios, 2014), are popular optimization tools. Yet in our setting, *neither the gradient* $\nabla L_{\text{SumVar}}(\boldsymbol{w})$ *nor its components can be computed exactly* and so estimations are required. To accurately estimate $\nabla L_{\text{SumVar}}(\boldsymbol{w}^{(t)})$ at round t, sufficient simulation steps are required, making the learning cost unaffordable and even exceed the cost of applying IS for each $\mathcal{E}_n$ individually. When the objectives are *non-smooth*, i.e., a pointwise maximum function $L(\boldsymbol{\sigma}(\boldsymbol{w})) \triangleq L_{\text{SimCos}}(\boldsymbol{w})$ with smooth components, the gradient mapping based method proposed in (Nesterov, 1998) guarantees an exponential regret convergence. However, when applying to our setting, it faces the same problem of expensive gradient (or its components) estimation. A more challenging point is, the updating of $\boldsymbol{w}^{(t)}$ has a fixed step size $\frac{1}{t}$ and constrained moving directions, i.e., $\boldsymbol{w}^{(t)} = \boldsymbol{w}^{(t-1)} + \frac{1}{t}(\boldsymbol{e}_{I_t} - \boldsymbol{w}^{(t-1)})$.

MIS-Learning vs. Multi-arm Bandit Optimization: The MIS-Learning is also similar to the bandit optimization problem (Hazan, 2012; Shalev-Shwartz et al., 2012), where at each round t, we pick an action $\boldsymbol{e}_{I_t}$ and observe information on the loss function L. The major difference is that these works consider a cumulative regret related to $\frac{1}{T}\sum_{t\in[T]} L(\boldsymbol{e}_{I_t})$ but we focus on the global loss $L(\frac{1}{T}\sum_{t\in[T]} \boldsymbol{e}_{I_t})$. Problems related to bandit optimization with global loss have been studied in (Rakhlin et al., 2011; Agrawal et al., 2016; Agrawal & Devanur, 2014; Berthet & Perchet, 2017). where



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

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

Figure 3: The reduction of cost measure $L_{\text{SumVar}}(\boldsymbol{x})$ (or $L_{\text{SimCos}}(\boldsymbol{x})$) 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 .

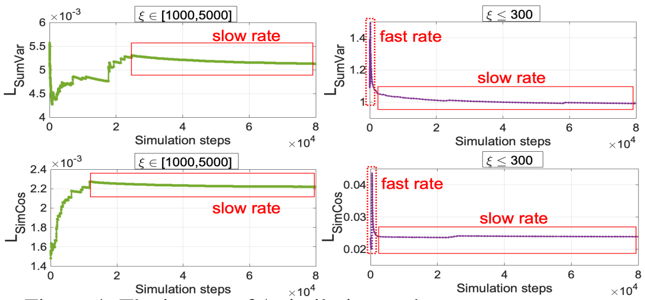


Figure 4: The impact of ξ-similarity on the convergence rate

they consider minimizing a known loss $L(\boldsymbol{w}^{(t)\top}V)$ with an unknown matrix V. This is different from our setting where L is unknown and cannot be compute analytically. (Agrawal & Devanur, 2014; Agrawal et al., 2016) consider a stochastic setting and achieve a convergence rate of $O(\sqrt{1/T})$. (Rakhlin et al., 2011) considers an adversarial setting, but there are cases that the regret cannot converge to zero. Our SumVar case is similar to (Berthet & Perchet, 2017), which considers the global loss $L(\frac{1}{T}\sum_{t\in[T]}\boldsymbol{e}_{I_t})$ and focuses on the strongly-convex and smooth loss function L. They consider $L(\boldsymbol{w})\triangleq\sum_{n\in[N]}\frac{1}{w_n}\sigma_n^2$, where $\sigma_n^2, n\in[N]$ are unknown but fixed. While in our setting, $\sigma_n^2, n\in[N]$ also depend on $\boldsymbol{w}$.

7. Conclusion

This paper develops an online learning (OL) framework to address the high simulation cost limitation of IS in dealing with *a set of rare events*. Our framework consists of a mixture importance sampling optimization problem (MISOP) and OL algorithms. MISOP aims to select the optimal mixture attaining various tradeoffs, which are quantified by our cost measures. We also show that the objective function is computationally expensive to evaluate, and extend MISOP to an OL setting to efficiently optimize the objective function without incurring any learning cost. Our OL algorithms learn the optimal mixture from simulation samples, and have the following regret: (1) learning to minimize sum of variances (SumVar MIS-Learning) with a regret $(\ln T)^2/T$; (2) learning to minimize the simulation cost (MinCos MIS-Learning) with a regret of $\sqrt{\ln T/T}$, where T denotes the number of samples. We demonstrate our methods on a realistic network and our methods reduce the cost measure value by 61.6%, compared with the uniform mixture IS.

References

Abilene. https://www.internet2.edu/.

Agrawal, S. and Devanur, N. R. Bandits with concave rewards and convex knapsacks. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pp. 989–1006. ACM, 2014.

Agrawal, S., Devanur, N. R., and Li, L. An efficient algorithm for contextual bandits with knapsacks, and an extension to concave objectives. In *Conference on Learning Theory*, pp. 4–18, 2016.

Berthet, Q. and Perchet, V. Fast rates for bandit optimization with upper-confidence frank-wolfe. In *Advances in Neural Information Processing Systems*, pp. 2225–2234, 2017.

Bucklew, J. *Introduction to rare event simulation*. Springer Science & Business Media, 2013.

Frank, J., Mannor, S., and Precup, D. Reinforcement learning in the presence of rare events. In *Proceedings of the 25th international conference on Machine learning*, pp. 336–343. ACM, 2008.

Frank, M. and Wolfe, P. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2): 95–110, 1956.

Hazan, E. The convex optimization approach to regret minimization. *Optimization for machine learning*, 2012.

Jain, S., Kumar, A., Mandal, S., Ong, J., Poutievski, L., Singh, A., Venkata, S., Wanderer, J., Zhou, J., Zhu, M., et al. B4: Experience with a globally-deployed software defined wan. In *ACM SIGCOMM Computer Communication Review*, volume 43, pp. 3–14. ACM, 2013.

Jiang, W., Zhang-Shen, R., Rexford, J., and Chiang, M. Cooperative content distribution and traffic engineering in an isp network. In *ACM SIGMETRICS Performance Evaluation Review*, volume 37, pp. 239–250. ACM, 2009.

Lacoste-Julien, S. and Jaggi, M. An affine invariant linear convergence analysis for frank-wolfe algorithms. *arXiv preprint arXiv:1312.7864*, 2013.

Liu, T. and Lui, J. C. S. FAVE: a fast and efficient network flow AVailability estimation method with bounded relative error. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, Apr. 2019.

Nesterov, Y. Introductory lectures on convex programming volume i: Basic course. *Lecture notes*, 1998.

Papamakarios, G. Comparison of modern stochastic optimization algorithms. 2014.

Rakhlin, A., Sridharan, K., and Tewari, A. Online learning: Beyond regret. 2011.

Rubino, G. and Tuffin, B. *Rare event simulation using Monte Carlo methods*. John Wiley & Sons, 2009.

Shalev-Shwartz, S. et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

Zhang, Y. Six months of Abilene traffic matrices. `http://www.cs.utexas.edu/ yzhang/`, 2014.

Appendix

A. Convexity and smoothness analysis

Theorem 2. *If $\{\mathcal{E}_n\}_{n=1}^N$ has a ξ-similarity, $L_{\mathrm{SumVar}}(\boldsymbol{w})$ given by Eq. (10) is α-strongly convex and β-smooth with:*

$$\alpha=\frac{2(\sum_{n\in[N]}\mu_n)^2}{N\xi^2} \text{ and } \beta=2\xi^3\sum_{n\in[N]}\mu_n. \tag{A.1}$$

Proof of Theorem 2: For the ease of presentation, in this proof, let $\boldsymbol{x}\sim Q$ represent $\boldsymbol{x}\sim Q(\boldsymbol{x};\boldsymbol{w})$ and $\boldsymbol{x}\sim P$ represent $\boldsymbol{x}\sim P(\boldsymbol{x})$. We have:

$$\begin{aligned}\sigma_n^2(\boldsymbol{w})&=\mathbb{V}_{\boldsymbol{x}\sim Q}\left[\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\frac{P(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}\right]\\&=\mathbb{E}_{\boldsymbol{x}\sim Q}\left[\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\frac{P^2(\boldsymbol{x})}{Q^2(\boldsymbol{x};\boldsymbol{w})}\right]-\mu_n^2\\&=\sum_{\boldsymbol{x}\in\Omega}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\frac{P^2(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}-\mu_n^2.\end{aligned} \tag{A.2}$$

Denote $\boldsymbol{Q}(\boldsymbol{x})=(Q_1(\boldsymbol{x}),...,Q_N(\boldsymbol{x}))$. The gradient of one-run variance σ_n^2 should be:

$$\begin{aligned}\nabla\sigma_n^2&=-\mathbb{E}_{\boldsymbol{x}\sim Q}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\frac{P^2(\boldsymbol{x})}{Q^3(\boldsymbol{x};\boldsymbol{w})}\boldsymbol{Q}(\boldsymbol{x})\\&=-\sum_{\boldsymbol{x}\in\Omega}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\frac{P^2(\boldsymbol{x})}{Q^2(\boldsymbol{x};\boldsymbol{w})}\boldsymbol{Q}(\boldsymbol{x})\end{aligned} \tag{A.3}$$

The gradient of $\nabla\left(\sum_{n\in[N]}\sigma_n^2\right)$, i.e., the Hessian matrix of $L_{\mathrm{SumVar}}(\boldsymbol{w})$ in Eq. (10) should be:

$$H(\boldsymbol{w})=2\sum_{\boldsymbol{x}\in\Omega}\frac{P^2(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q^3(\boldsymbol{x};\boldsymbol{w})}\boldsymbol{Q}(\boldsymbol{x})\boldsymbol{Q}(\boldsymbol{x})^\top \tag{A.4}$$

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

$$\boldsymbol{Q}(\boldsymbol{x})\boldsymbol{Q}(\boldsymbol{x})^\top\succeq 0\Rightarrow H(\boldsymbol{w})\succeq 0. \tag{A.5}$$

(2) About the *α-strongly convexity*,

$$H(\boldsymbol{w})\succeq\alpha I$$

$$\Leftrightarrow\forall\boldsymbol{z}\in\Delta,\boldsymbol{z}^\top H(w)\boldsymbol{z}\geq\alpha\boldsymbol{z}^\top\boldsymbol{z}=\alpha; \tag{A.6}$$

$$\Leftrightarrow\forall\boldsymbol{z}\in\Delta,\sum_{\boldsymbol{x}\in\Omega}\frac{P^2(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q^3(\boldsymbol{x};\boldsymbol{w})}Q^2(\boldsymbol{x};\boldsymbol{z})\geq\frac{\alpha}{2}; \tag{A.7}$$

$$\Leftrightarrow\forall\boldsymbol{z}\in\Delta,\mathop{\mathbb{E}}_{\boldsymbol{x}\sim P}\frac{P(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}\left\{\frac{Q(\boldsymbol{x};\boldsymbol{z})}{Q(\boldsymbol{x};\boldsymbol{w})}\right\}^2\geq\frac{\alpha}{2}. \tag{A.8}$$

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

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{x}\sim P}\frac{P(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}\left\{\frac{Q(\boldsymbol{x};\boldsymbol{z})}{Q(\boldsymbol{x};\boldsymbol{w})}\right\}^2\\&\geq\frac{1}{\xi^2}\mathbb{E}_{\boldsymbol{x}\sim P}\frac{P(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}.\end{aligned} \tag{A.9}$$

According to the Cauchy-Schwarz inequality,

$$\begin{aligned}&\sum_{\boldsymbol{x}\in\Delta}\frac{P^2(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}\sum_{\boldsymbol{x}\in\Delta}\left[Q(\boldsymbol{x};\boldsymbol{w})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\right]\\&\geq\left[\sum_{\boldsymbol{x}\in\Delta}\left(P(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\right)\right]^2\\&=\left(\mathop{\mathbb{E}}_{\boldsymbol{x}\sim P}\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\right)^2=\left(\sum_{n\in[N]}\mu_n\right)^2.\end{aligned} \tag{A.10}$$

Note that:

$$\sum_{\boldsymbol{x}\in\Delta}\left[Q(\boldsymbol{x};\boldsymbol{w})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\right]=\mathop{\mathbb{E}}_{\boldsymbol{x}\sim Q}\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\leq N, \tag{A.11}$$

we have the following:

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{x}\sim P}\frac{P(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}=\sum_{\boldsymbol{x}\in\Delta}\frac{P^2(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}\\&\geq\frac{(\sum_{n\in[N]}\mu_n)^2}{\sum_{\boldsymbol{x}\in\Delta}Q(\boldsymbol{x};\boldsymbol{w})\sum_{i\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}=\frac{(\sum_{n\in[N]}\mu_n)^2}{N}\end{aligned} \tag{A.12}$$

Hence, we can take α as:

$$\alpha=\frac{2(\sum_{n\in[N]}\mu_i)^2}{N\xi^2} \tag{A.13}$$

(3) About the *β-smoothness*,

$$H(\boldsymbol{w})\preceq\beta I$$

$$\Leftrightarrow\forall\boldsymbol{z}\in\Delta,\boldsymbol{z}^\top H(w)\boldsymbol{z}\leq\beta\boldsymbol{z}^\top\boldsymbol{z}=\beta; \tag{A.14}$$

$$\Leftrightarrow\forall\boldsymbol{z}\in\Delta,\sum_{\boldsymbol{x}\in\Omega}\frac{P^2(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q^3(\boldsymbol{x};\boldsymbol{w})}Q^2(\boldsymbol{x};\boldsymbol{z})\leq\frac{\beta}{2}; \tag{A.15}$$

$$\Leftrightarrow\forall\boldsymbol{z}\in\Delta,\mathop{\mathbb{E}}_{\boldsymbol{x}\sim P}\frac{P(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}\left\{\frac{Q(\boldsymbol{x};\boldsymbol{z})}{Q(\boldsymbol{x};\boldsymbol{w})}\right\}^2\leq\frac{\beta}{2}. \tag{A.16}$$

Still, by the definition of ξ-similarity, we have:

$$\begin{aligned}&\mathop{\mathbb{E}}_{\boldsymbol{x}\sim P}\frac{P(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}\left\{\frac{Q(\boldsymbol{x};\boldsymbol{z})}{Q(\boldsymbol{x};\boldsymbol{w})}\right\}^2\\&\leq\xi^2\mathop{\mathbb{E}}_{\boldsymbol{x}\sim P}\frac{P(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}t\leq\xi^2\sum_{n\in[N]}\mathop{\mathbb{E}}_{\boldsymbol{x}\sim P}\frac{P(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}.\end{aligned} \tag{A.17}$$

As $Q_n(\boldsymbol{x})$ is the "customized" IS distribution of $\mathcal{E}_n$, it simulates the occurrence of $\mathcal{E}_n$ more often than the natural distribution $P(\boldsymbol{x})$, i.e., $Q_n(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\geq P(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})$. Hence, when $\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})=1$,

$$\begin{aligned}Q(\boldsymbol{x};\boldsymbol{w})&\geq w_nQ_n(\boldsymbol{x})+(1-w_n)\tfrac{1}{\xi}Q_n(\boldsymbol{x})\\&\geq\left(w_n+\frac{(1-w_n)}{\xi}\right)P(\boldsymbol{x})\geq\frac{1}{\xi}P(\boldsymbol{x}).\end{aligned} \tag{A.18}$$

$$\frac{P(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\leq\xi\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x}). \tag{A.19}$$

$$\mathbb{E}_{\boldsymbol{x}\sim P}\frac{P(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}\leq\xi\mathbb{E}_{\boldsymbol{x}\sim P}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})=\xi\mu_n. \tag{A.20}$$

Therefore,

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim P}\frac{P(\boldsymbol{x})\sum_{n\in[N]}\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w})}\left\{\frac{Q(\boldsymbol{x};\boldsymbol{z})}{Q(\boldsymbol{x};\boldsymbol{w})}\right\}^2\leq\xi^3\sum_{n\in[N]}\mu_n\leq\frac{\beta}{2}. \tag{A.21}$$

We can take β as:

$$\beta=2\xi^3\sum_{n\in[N]}\mu_n. \tag{A.22}$$

Then, by the definition of Theorem 2, we finish the proof. ∎

Corollary 1. *If $\{\mathcal{E}_n\}_{n=1}^N$ has a ξ-similarity, $\ell_n(\boldsymbol{w}),n\in[N]$ in Eq.(9) is α_n-strongly convex and β_n-smooth, where*

$$\alpha_n=\frac{2\mu_n^2}{\xi^2(\mu_n-o_n)^2} \text{ and } \beta_n=\frac{2\xi^3\mu_n}{(\mu_n-o_n)^2}. \tag{A.23}$$

Proof of Corollary 1: This can be considered as a special case of Theorem 2 when $\{\mathcal{E}_n\}_{n=1}^N$ contains only a single event. The proof follows the same way as Theorem 2. ∎

Corollary 2. *If $\{\mathcal{E}_n\}_{n=1}^N$ has a ξ-similarity, $L_{\mathrm{SimCos}}(\boldsymbol{w})$ in Equation (12) is α'-strongly convex, where*

$$\alpha'\triangleq\min_{n\in[N]}\alpha_n=\min_{n\in[N]}\frac{2\mu_n^2}{\xi^2(\mu_n-o_n)^2}. \tag{A.24}$$

Proof of Corollary 2: The proof can be finished by simply using the strong convexity of $\ell_n(\boldsymbol{w})$ (i.e., the components of $L_{\mathrm{SimCos}}(\boldsymbol{w})$) in Corollary 1, and the pointwise maximum property of $L_{\mathrm{SimCos}}(\boldsymbol{w})$. ∎

B. Regret analysis for SumVar MIS-Learning.

We provide the complete analysis for the regret upper bound of SumVar algorithm.

Let I_t stand for the index of arm selected by $\mathcal{A}$ at round t. Recall that $\boldsymbol{w}^{(t)}$ is defined as the proportions of arm selections, i.e., $\boldsymbol{w}^{(t)}=\frac{1}{t}\sum_{s\in[t]}\boldsymbol{e}_{I_s}$. We can derive the following recurrence:

$$\boldsymbol{w}^{(t+1)}=\frac{t\boldsymbol{w}^{(t)}+\boldsymbol{e}_{I_{t+1}}}{t+1}=\boldsymbol{w}^{(t)}+\frac{\boldsymbol{e}_{I_{t+1}}-\boldsymbol{w}^{(t)}}{t+1}. \tag{B.1}$$

Let $\boldsymbol{w}^*$ be the optimal mixture:

$$\boldsymbol{w}^*=\operatorname{argmin}_{\boldsymbol{w}\in\Delta} L_{\mathrm{SumVar}}(\boldsymbol{w}), \tag{B.2}$$

and define $\boldsymbol{e}_{*_{t+1}}$ as the following minimizer:

$$\boldsymbol{e}_{*_{t+1}}=\operatorname{argmin}_{\boldsymbol{z}\in\Delta} \boldsymbol{z}^\top\nabla L_{\mathrm{SumVar}}(\boldsymbol{w}^{(t)}), \tag{B.3}$$

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

$$\varepsilon^{(t+1)}=\nabla L_{\mathrm{SumVar}}(\boldsymbol{w}^{(t)})^\top(\boldsymbol{e}_{I_{t+1}}-\boldsymbol{e}_{*_{t+1}}). \tag{B.4}$$

The regret analysis of the SumVar algorithm can be broken down into the following steps.

Step 1: We first partition the regret R_T and show that

$$R_T=\frac{1}{T}\Big[\sum_{s\in[T]}\frac{\beta}{s}+\sum_{s\in[T]}\varepsilon^{(s)}\Big]\le\beta\frac{\ln T}{T}+\frac{\sum_{s\in[T]}\varepsilon^{(s)}}{T}. \tag{B.5}$$

We claim the following recurrent relationship:

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

Proof of Claim B.1: By the convexity and β-smoothness properties of $L_{\mathrm{SumVar}}(\boldsymbol{w})$, we have:

$$\begin{aligned}
&L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t+1)}\big)-L_{\mathrm{SumVar}}(\boldsymbol{w}^*)\\
=&L_{\mathrm{SumVar}}\Big(\boldsymbol{w}^{(t)}+\tfrac{1}{t+1}\big(\boldsymbol{e}_{I_{t+1}}-\boldsymbol{w}^{(t)}\big)\Big)-L_{\mathrm{SumVar}}(\boldsymbol{w}^*)\\
\le&L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)+\tfrac{1}{t+1}\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)^\top\big(\boldsymbol{e}_{I_{t+1}}-\boldsymbol{w}^{(t)}\big)\\
&+\tfrac{\beta}{2(t+1)^2}\left\|\boldsymbol{e}_{I_{t+1}}-\boldsymbol{w}^{(t)}\right\|_2^2-L_{\mathrm{SumVar}}(\boldsymbol{w}^*)\\
=&\big[L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)-L_{\mathrm{SumVar}}(\boldsymbol{w}^*)\big]\\
&+\tfrac{1}{t+1}\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)^\top\big(\boldsymbol{e}_{I_{t+1}}-\boldsymbol{e}_{*_{t+1}}\big)\\
&+\tfrac{1}{t+1}\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)^\top\big(\boldsymbol{e}_{*_{t+1}}-\boldsymbol{w}^{(t)}\big)+\tfrac{\beta}{2(t+1)^2}\left\|\boldsymbol{e}_{I_{t+1}}-\boldsymbol{w}^{(t)}\right\|_2^2\\
\le&\big[L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)-L_{\mathrm{SumVar}}(\boldsymbol{w}^*)\big]\\
&+\tfrac{1}{t+1}\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)^\top\big(\boldsymbol{e}_{I_{t+1}}-\boldsymbol{e}_{*_{t+1}}\big)\\
&+\tfrac{1}{t+1}\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)^\top\big(\boldsymbol{e}_{*_{t+1}}-\boldsymbol{w}^{(t)}\big)+\tfrac{\beta}{(t+1)^2}\\
\le&\big[L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)-L_{\mathrm{SumVar}}(\boldsymbol{w}^*)\big]\\
&+\tfrac{1}{t+1}\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)^\top\big(\boldsymbol{e}_{I_{t+1}}-\boldsymbol{e}_{*_{t+1}}\big)\\
&+\tfrac{1}{t+1}\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)^\top\big(\boldsymbol{w}^*-\boldsymbol{w}^{(t)}\big)+\tfrac{\beta}{(t+1)^2}\\
\le&\tfrac{t}{t+1}\big[L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)-L_{\mathrm{SumVar}}(\boldsymbol{w}^*)\big]\\
&+\tfrac{1}{t+1}\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)^\top\big(\boldsymbol{e}_{I_{t+1}}-\boldsymbol{e}_{*_{t+1}}\big)+\tfrac{\beta}{(t+1)^2}.
\end{aligned} \tag{B.6}$$

By combining Eq. (B.4) and Eq. (B.6), we finish the proof of Claim B.1. ■

Resolve the recurrent in Claim B.1 we can obtain:

$$(t+1)R_{t+1}\le\sum_{s\in[t+1]}\frac{\beta}{s}+\sum_{s\in[t+1]}\varepsilon^{(s)}. \tag{B.7}$$

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

Step 2: To bound R_T, we consider bounding $\sum_{s\in[T]}\varepsilon^{(s)}$.

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

Claim 1. *Assume $c_n^{(t)}$ satisfies*

$$\mathbb{P}\Big[g_n^{(t)}-\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)\big|_n\le c_n^{(t)}\Big]\le\zeta^{(t)}, \tag{B.8}$$

$$\mathbb{P}\Big[g_n^{(t)}-\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)\big|_n\ge -c_n^{(t)}\Big]\le\zeta^{(t)}. \tag{B.9}$$

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

Proof of Claim 1: According to the union bound, we know with a probability at least $1-2\zeta^{(t)}$, we have:

$$\begin{aligned}
\varepsilon^{(t+1)}=&\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)^\top\big(\boldsymbol{e}_{I_{t+1}}-\boldsymbol{e}_{*_{t+1}}\big)\\
=&\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)\big|_{I_{t+1}}-\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)\big|_{*_{t+1}}\\
\le&\big(g_{I_{t+1}}^{(t)}+c_{I_{t+1}}^{(t)}\big)-\big(g_{*_{t+1}}^{(t)}-c_{*_{t+1}}^{(t)}\big)\\
\le&\big(g_{I_{t+1}}^{(t)}+c_{I_{t+1}}^{(t)}\big)-\big(g_{I_{t+1}}^{(t)}-c_{I_{t+1}}^{(t)}\big)\\
=&2c_{I_{t+1}}^{(t)}
\end{aligned} \tag{B.10}$$

We finish the proof of Claim 1. ■

We then derive the expression of $c_n^{(t)}$.

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

$$\varepsilon_n^{(t)}=\frac{\ln\frac{1}{\zeta^{(t)}}}{3t}Z_n^{max}+\sqrt{\frac{1}{9t^2}(\ln\frac{1}{\zeta^{(t)}}Z_n^{max})^2+\frac{2}{t}\ln\frac{1}{\zeta^{(t)}}\mathbb{V}Z_n(\boldsymbol{x})}.$$

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

$$\mathbb{P}\Big[g_n^{(t)}-\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)\big|_n\le\epsilon_n^{(t)}\Big]\le\zeta^{(t)}, \tag{B.11}$$

$$\mathbb{P}\Big[g_n^{(t)}-\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)\big|_n\ge-\epsilon_n^{(t)}\Big]\le\zeta^{(t)}. \tag{B.12}$$

Proof of Theorem 3: We utilize the *Bernstein Inequality* to finish the proof. Recall that $g_n^{(t)}=\frac{-1}{t-1}\sum_{s\in[t-1]}Z_n(\boldsymbol{x}^{(s)})$ and $\nabla L_{\mathrm{SumVar}}\big(\boldsymbol{w}^{(t)}\big)\big|_n=-\mathbb{E}\,Z_n(\boldsymbol{x})$. Then by the definition of ξ-similarity and $P(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})\le Q_n(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})$, we have:

$$\begin{aligned}
\frac{P(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q(\boldsymbol{x};\boldsymbol{w}^{(t-1)})}\le&\frac{Q_n(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{w_n^{(t-1)}Q_i(\boldsymbol{x})+\sum_{i\neq n}w_i^{(t-1)}Q_i(\boldsymbol{x})}\\
\le&\frac{Q_n(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{w_n^{(t-1)}Q_n(\boldsymbol{x})+\big(1-w_n^{(t-1)}\big)\frac{1}{\xi}Q_n(\boldsymbol{x})}\\
\le&\xi\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x}).
\end{aligned} \tag{B.13}$$

Therefore, we obtain the bound of $Z_n(\boldsymbol{x})$ by the following:

$$Z_n(\boldsymbol{x})=\sum_{i\in[N]}\frac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})Q_n(\boldsymbol{x})}{Q^3(\boldsymbol{x};\boldsymbol{w}^{(t-1)})}\le\sum_{i\in[N]}\frac{Q_n^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})Q_n(\boldsymbol{x})}{Q^3(\boldsymbol{x};\boldsymbol{w}^{(t-1)})}$$

$$\leq \xi^3 \sum_{i\in[N]} \mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x}) \leq N\xi^3. \tag{B.14}$$

Similarly, we can obtain the bounds of $\mathbb{E}Z_n(\boldsymbol{x})$ and $\mathbb{E}\,Z_n^2(\boldsymbol{x})$:

$$\begin{aligned}\mathbb{E}Z_n(\boldsymbol{x}) &= \sum_{i\in[N]} \sum_{\boldsymbol{x}\in\Omega} \frac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})Q_n(\boldsymbol{x})}{Q^2(\boldsymbol{x};\boldsymbol{w}^{(t-1)})} \\ &\leq \sum_{i\in[N]} \sum_{\boldsymbol{x}\in\Omega} \frac{P(\boldsymbol{x})Q_i(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})Q_n(\boldsymbol{x})}{Q^2(\boldsymbol{x};\boldsymbol{w}^{(t-1)})} \\ &\leq \xi^2 \sum_{i\in[N]} \sum_{\boldsymbol{x}\in\Omega} \mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})P(\boldsymbol{x}) \\ &= \xi^2 \sum_{i\in[N]} \mu_i \end{aligned}\tag{B.15}$$

$$\begin{aligned}\mathbb{E}Z_n^2(\boldsymbol{x}) &= \sum_{\boldsymbol{x}\in\Omega} \frac{P^4(\boldsymbol{x})\left(\sum_{i\in[N]}\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})\right)^2 Q_n^2(\boldsymbol{x})}{Q^5(\boldsymbol{x};\boldsymbol{w}^{(t-1)})} \\ &\leq \sum_{i,j\in[N]} \sum_{\boldsymbol{x}\in\Omega} \frac{P^4(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_j}(\boldsymbol{x})Q_n^2(\boldsymbol{x})}{Q^5(\boldsymbol{x};\boldsymbol{w}^{(t-1)})} \\ &\leq \sum_{i,j\in[N]} \sum_{\boldsymbol{x}\in\Omega} \frac{P(\boldsymbol{x})Q_i^2(\boldsymbol{x})Q_j(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_j}(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_j}(\boldsymbol{x})Q_n^2(\boldsymbol{x})}{Q^5(\boldsymbol{x};\boldsymbol{w}^{(t-1)})} \\ &\leq \xi^5 \sum_{i,j\in[N]} \sum_{\boldsymbol{x}\in\Omega} \mathbf{1}_{\mathcal{E}_j}(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_j}(\boldsymbol{x})P(\boldsymbol{x}) \\ &= \xi^5 N \sum_{i\in[N]} \sum_{\boldsymbol{x}\in\Omega} \mathbf{1}_{\mathcal{E}_j}(\boldsymbol{x})P(\boldsymbol{x}) \\ &= N\xi^5 \sum_{i\in[N]} \mu_i \end{aligned}\tag{B.16}$$

By centralizing $Z_n(\boldsymbol{x})$, we have:

$$\mathbb{V}[Z_n(\boldsymbol{x}) - \mathbb{E}\,Z_n(\boldsymbol{y})] \leq N\xi^5 \sum_{i\in[N]} \mu_i \tag{B.17}$$

$$\begin{aligned}|Z_n(\boldsymbol{x}) - \mathbb{E}\,Z_n(\boldsymbol{y})| &\leq \max\{\xi^3 \sum_{i\in[N]} \mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x}), \xi^2 \sum_{i\in[N]} \mu_i\} \\ &\leq N\xi^3 \end{aligned}\tag{B.18}$$

Hence, $Z_n^{\max} \leq N\xi^3$. Then by the Bernstein inequality:

$$\begin{aligned}&\mathbb{P}\left[\tfrac{1}{t}\sum_{s\in[t]} Z_n(\boldsymbol{x}^{(s)}) - \mathbb{E}Z_n(\boldsymbol{y}) \geq \epsilon_n^{(t)}\right] \leq e^{-\frac{t\left(\epsilon_n^{(t)}\right)^2}{2\mathbb{V}Z_n + \frac{2}{3}Z_n^{\max}\epsilon_n^{(t)}}}, \\ &\mathbb{P}\left[\tfrac{1}{t}\sum_{s\in[t]} Z_n\left(\boldsymbol{x}^{(s)}\right) - \mathbb{E}Z_n(\boldsymbol{y}) \leq -\epsilon_n^{(t)}\right] \leq e^{-\frac{t\left(\epsilon_n^{(t)}\right)^2}{2\mathbb{V}Z_n + \frac{2}{3}Z_n^{\max}\epsilon_n^{(t)}}}. \end{aligned}\tag{B.19}$$

By solving:

$$e^{-\frac{t\left(\epsilon_n^{(t)}\right)^2}{2\mathbb{V}Z_n + \frac{2}{3}Z_n^{\max}\epsilon_n^{(t)}}} = \zeta^{(t)}, \tag{B.20}$$

we obtain the expression of ϵ_n.

The proof of Theorem 3 is done. ∎

We then derive the expression of $c_n^{(t)}$. By Theorem 3, we know that Eq. (B.8) and (B.9) can be satisfied with the choice of $c_n^{(t)} = \epsilon_n^{(t)}$ where $\epsilon_n^{(t)}$ is defined in Theorem 3.

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

Claim 2. *With the choice of $c_n^{(t)} = \epsilon_n^{(t)}$, we have:*

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

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

Proof of Claim 2: By plugging Eq. (B.17) and (B.18) into the expression of $\epsilon_n^{(t)}$ given in Theorem 3, we have:

$$\begin{aligned}c_n^{(t)} \leq &\sqrt{\tfrac{1}{9}N^2\xi^6\left(\frac{\ln\frac{1}{\zeta^{(t)}}}{t}\right)^2 + 2N\xi^5 \sum_{i\in[N]}\mu_i \frac{\ln\frac{1}{\zeta^{(t)}}}{t}} \\ &+ \tfrac{1}{3}N\xi^3 \frac{\ln\frac{1}{\zeta^{(t)}}}{t}. \end{aligned}\tag{B.21}$$

If $\frac{\ln\frac{1}{\zeta^{(t)}}}{t} \geq \frac{9\sum_{m\in[N]}\mu_m}{4N\xi}$, then

$$2N\xi^5 \sum_{i\in[N]}\mu_i \frac{\ln\frac{1}{\zeta^{(t)}}}{t} \leq \tfrac{8}{9}N^2\xi^6\left(\frac{\ln\frac{1}{\zeta^{(t)}}}{t}\right)^2, \tag{B.22}$$

$$\sqrt{\tfrac{1}{9}N^2\xi^6\left(\frac{\ln\frac{1}{\zeta^{(t)}}}{t}\right)^2 + 2N\xi^5 \sum_{i\in[N]}\mu_i \frac{\ln\frac{1}{\zeta^{(t)}}}{t}} \leq N\xi^3 \frac{\ln\frac{1}{\zeta^{(t)}}}{t}. \tag{B.23}$$

Similarly, we can show the case for $\frac{\ln\frac{1}{\zeta^{(t)}}}{t} < \frac{9\sum_{i\in[N]}\mu_i}{4N\xi}$. ∎

Remark: Combine Claim 1 and 2, we find that, when t is small, we have $\frac{\ln\frac{1}{\zeta^{(t)}}}{t} \geq \frac{9\sum_{i\in[N]}\mu_i}{4N\xi}$ and so $\varepsilon^{(t)}$ converges at a fast speed of $O\left(\frac{1}{t}\ln\frac{1}{\zeta^{(t)}}\right)$. As t increases, the convergence speed of $\varepsilon^{(t)}$ slows down to $O\left(\sqrt{\frac{1}{t}\ln\frac{1}{\zeta^{(t)}}}\right)$. Combine Claim 1, 2 and Theorem 3, we show that with the choice of $c_n^{(t)}$ in Theorem 3, 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 2.

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

By directly applying the conclusion in (Berthet & Perchet, 2017; Lacoste-Julien & Jaggi, 2013), we have:

Claim B.2. *Denote η as the distance from $\boldsymbol{w}^*$ to $\partial\Delta^N$. Then $\beta \geq \alpha\eta^2$ and*

$$\begin{aligned}&\nabla L_{\mathrm{SumVar}}\left(\boldsymbol{w}^{(t)}\right)^\top\left(\boldsymbol{w}^{(t)} - e_{*t+1}\right) \\ &\geq \sqrt{2\alpha\eta^2}\sqrt{L_{\mathrm{SumVar}}\left(\boldsymbol{w}^{(t)}\right) - L_{\mathrm{SumVar}}\left(\boldsymbol{w}^*\right)}. \end{aligned}\tag{B.24}$$

Plug Eq. (B.24) into Eq. (B.6) we will have:

$$\begin{aligned}R_{t+1} &\leq R_t + \tfrac{\varepsilon^{(t+1)}}{t+1} + \tfrac{1}{t+1}\nabla L_{\mathrm{SumVar}}\left(\boldsymbol{w}^{(t)}\right)^\top\left(\boldsymbol{e}_{*_{t+1}} - \boldsymbol{w}^{(t)}\right) + \tfrac{\beta}{(t+1)^2} \\ &\leq R_t + \tfrac{\varepsilon^{(t+1)}}{t+1} - \tfrac{\sqrt{2\alpha\eta^2}}{t+1}\sqrt{R_t} + \tfrac{\beta}{(t+1)^2}. \end{aligned}\tag{B.25}$$

Denote $\psi(x) = x^2 - \sqrt{2\alpha\eta^2}x$. We have:

$$\begin{aligned}(t+1)R_{t+1} \leq &tR_t + \frac{\left(\varepsilon^{(t+1)}\right)^2}{2\alpha\eta^2} + \frac{\beta}{t+1} \\ &+ \left[\psi\left(\sqrt{R_t}\right) - \psi\left(\tfrac{\varepsilon^{(t+1)}}{\sqrt{2\alpha\eta^2}}\right)\right]. \end{aligned}\tag{B.26}$$

Let $c^{(t)} \triangleq \max_{n\in[N]} c_n^{(t)}$. Claim 2 provides an upper bound of $c_n^{(t)}$. Next, we utilize $c^{(t)}$ to bound R_T:

Claim 3. *Assume we select $\zeta^{(t)}$ properly such that*

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

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

$$\begin{aligned}TR_T \leq &\tfrac{\alpha\eta^2}{2} + \tfrac{\pi^2\beta^2}{3\alpha\eta^2} + \beta\ln T \\ &+ \tfrac{8\beta}{\alpha\eta^2}\sum_{t\in[T]}\tfrac{c^{(t)}}{t} + \tfrac{8}{\alpha\eta^2}\sum_{t\in[T]}\left(c^{(t)}\right)^2. \end{aligned}\tag{B.27}$$

Proof of Claim 3: When $R_T \geq \frac{\alpha\eta^2}{2}$, by Eq. (B.5) we have:

$$T \leq \tfrac{2}{\alpha\eta^2}\left(\sum_{s\in[T]}\tfrac{\beta}{s} + \varepsilon^{(s)}\right). \tag{B.28}$$

Therefore, the Caucy-Schwarz inequality implies:

$$\Big(\sum_{s\in[T]}\frac{\beta}{s}+\varepsilon^{(s)}\Big)^2\leq T\sum_{s\in[T]}\Big(\frac{\beta}{s}+\varepsilon^{(s)}\Big)^2$$
$$\leq\frac{2}{\alpha\eta^2}\sum_{s\in[T]}\Big(\frac{\beta}{s}+\varepsilon^{(s)}\Big)\sum_{s\in[T]}\Big(\frac{\beta}{s}+\varepsilon^{(s)}\Big)^2. \quad (B.29)$$

$$TR_T\leq\sum_{s\in[T]}\frac{\beta}{s}+\varepsilon^{(s)}\leq\frac{2}{\alpha\eta^2}\sum_{s\in[T]}\Big(\frac{\beta}{s}+\varepsilon^{(s)}\Big)^2$$
$$=\frac{2\beta^2}{\alpha\eta^2}\sum_{s\in[T]}\frac{1}{s^2}+\frac{4\beta}{\alpha\eta^2}\sum_{s\in[T]}\frac{\varepsilon^{(s)}}{s}+\frac{2}{\alpha\eta^2}\sum_{s\in[T]}\Big(\varepsilon^{(s)}\Big)^2 \quad (B.30)$$

By Claim 1, with a probability at least $1-N\sum_t\zeta^{(t)}$, we have:

$$TR_t\leq\frac{2\beta^2}{\alpha\eta^2}\sum_{s\in[T]}\frac{1}{s^2}+\frac{8\beta}{\alpha\eta^2}\sum_{s\in[T]}\frac{c^{(s)}}{s}+\frac{8}{\alpha\eta^2}\sum_{s\in[T]}(c^{(s)})^2$$
$$\leq\frac{\pi^2\beta^2}{3\alpha\eta^2}+\frac{8\beta}{\alpha\eta^2}\sum_{s\in[T]}\frac{c^{(s)}}{s}+\frac{8}{\alpha\eta^2}\sum_{s\in[T]}\Big(c^{(s)}\Big)^2. \quad (B.31)$$

Next, we consider the case that $R_T\leq\frac{\alpha\eta^2}{2}$. Denote T_0 as the last time before T where $R_t\geq\frac{\alpha\eta^2}{2}$. When $\varepsilon^{(t)}\leq 2c^{(t)}$ with a probability at least $1-\zeta^{(t)}$, by Eq. (B.26), we have

$$(t+1)R_{t+1}\leq tR_t+\frac{2(c^{(t+1)})^2}{\alpha\eta^2}+\frac{\beta}{(t+1)}$$
$$+\Big[\psi\big(\sqrt{R_t}\big)-\psi\Big(\frac{\sqrt{2}c^{(t+1)}}{\sqrt{\alpha\eta^2}}\Big)\Big]. \quad (B.32)$$

Therefore, for any $t>T_0$ such that $R_t\leq\frac{\alpha\eta^2}{2}$, if assume that $R_t\geq\frac{2}{\alpha\eta^2}\frac{\sum_{s\in[t]}(c^{(t)})^2}{t}\geq\frac{2}{\alpha\eta^2}(c^{(t+1)})^2$, then we have:

$$\psi\left(\sqrt{R_t}\right)-\psi\Big(\frac{\sqrt{2}c^{(t+1)}}{\sqrt{\alpha\eta^2}}\Big)\leq 0,$$

$$(t+1)R_{t+1}\leq tR_t+\frac{2\big(c^{(t+1)}\big)^2}{\alpha\eta^2}+\frac{\beta}{t+1}. \quad (B.33)$$

We then denote T_1 as the last time before T where we have $R_t<\frac{2}{\alpha\eta^2}\frac{\sum_{s\in[t]}(c^{(t)})^2}{t}$. If $T_1\geq T_0$, by applying the conclusion in Eq. (B.33), we have:

$$TR_T\leq(T_1+1)R_{T_1+1}+\sum_{s=T_1+2}^{T}\Big\{\frac{2\big(c^{(s)}\big)^2}{\alpha\eta^2}+\frac{\beta}{s}\Big\}. \quad (B.34)$$

Combine Claim B.1 and Claim 1, we have:

$$(T_1+1)R_{T_1+1}\leq T_1R_{T_1}+\frac{\beta}{(T_1+1)}+2c^{(T_1+1)} \quad (B.35)$$

Then by the definition of T_1:

$$T_1R_{T_1}<\sum_{s\in[T_1]}\frac{2(c^{(s)})^2}{\alpha\eta^2}. \quad (B.36)$$

Combine Eq. (B.34), (B.35) and (B.36), we show that with a probability at least $1-N\sum_t\zeta^{(t)}$:

$$TR_T\leq\sum_{s\in[T]}^{T}\Big\{\frac{2(c^{(s)})^2}{\alpha\eta^2}+\frac{\beta}{s}\Big\}+2c^{(T_1+1)}-\frac{2(c^{(T_1+1)})^2}{\alpha\eta^2}$$
$$\leq\frac{2}{\alpha\eta^2}\sum_{s\in[T]}^{T}(c^{(s)})^2+\beta\ln T+\frac{\alpha\eta^2}{2}. \quad (B.37)$$

If $T_1<T_0$, similar to the above analysis, we have:

$$TR_T\leq(T_0+1)R_{T_0+1}+\sum_{s=T_0+2}^{T}\Big\{\frac{2(c^{(s)})^2}{\alpha\eta^2}+\frac{\beta}{s}\Big\}$$
$$\leq T_0R_{T_0}+\frac{\beta}{(T_0+1)}+2c^{(T_0+1)}+\sum_{s=T_0+2}^{T}\Big\{\frac{2(c^{(s)})^2}{\alpha\eta^2}+\frac{\beta}{s}\Big\}$$
$$\leq T_0R_{T_0}+\sum_{s=T_0+1}^{T}\Big\{\frac{2(c^{(s)})^2}{\alpha\eta^2}+\frac{\beta}{s}\Big\}+\frac{\alpha\eta^2}{2}. \quad (B.38)$$

As $R_{T_0}\geq\frac{\alpha\eta^2}{2}$, by Eq. (B.31), we have:

$$T_0R_{T_0}\leq\frac{\pi^2\beta^2}{3\alpha\eta^2}+\frac{8\beta}{\alpha\eta^2}\sum_{s\in[T_0]}\frac{c^{(s)}}{s}+\frac{8}{\alpha\eta^2}\sum_{s\in[T_0]}\Big(c^{(s)}\Big)^2. \quad (B.39)$$

Hence, when $T_1<T_0$, with probability at least $1-N\sum_t\zeta^{(t)}$, we have:

$$TR_T\leq\beta\ln T+\frac{\alpha\eta^2}{2}+\frac{\pi^2\beta^2}{3\alpha\eta^2}$$
$$+\frac{8\beta}{\alpha\eta^2}\sum_{s\in[T]}\frac{c^{(s)}}{s}+\frac{8}{\alpha\eta^2}\sum_{s\in[T]}\Big(c^{(s)}\Big)^2. \quad (B.40)$$

By taking the maximum out of the upper bounds in Eq. (B.31), (B.37) and (B.40), we can prove Claim 3. ∎

Step 4: We now discuss how to select $\zeta^{(t)}$ to guarantee that $\frac{1}{t}\sum_{s\in[t]}\big(c^{(t)}\big)^2\geq\big(c^{(t+1)}\big)^2$, and give bounds of $\sum_t\big(c^{(t)}\big)^2$ and $\sum_t\frac{c^{(t)}}{t}$.

From the previous discussion, we know that:

$$c^{(t)}=\begin{cases}\frac{4}{3}N\xi^3\frac{\ln\frac{1}{\zeta^{(t)}}}{t}, & \text{if } \frac{\ln\frac{1}{\zeta^{(t)}}}{t}\geq\frac{9\sum_{i\in[N]}\mu_i}{4N\xi};\\ 2\sqrt{N\xi^5\sum_{i\in[N]}\mu_i\frac{\ln\frac{1}{\zeta^{(t)}}}{t}}, & \text{otherwise.}\end{cases}$$

Observe that $\frac{\sum_{s\in[t]}\big(c^{(t)}\big)^2}{t}\geq\big(c^{(t+1)}\big)^2$ is achieved if $c^{(t)}$ decreases with the increasing t. Also consider that we need to guarantee that $1-N\sum_t\zeta^{(t)}$ is large enough. We select $\zeta^{(t)}$ and bound $\sum_t\big(c^{(t)}\big)^2$ and $\sum_t\frac{c^{(t)}}{t}$ by the following:

Claim 4. *With the choice of* $\zeta^{(t)}=\begin{cases}T_0^{-2}, & \text{if } t\leq T_0;\\ t^{-2}, & \text{if } t>T_0.\end{cases}$

If $c^{(t)}=\frac{4}{3}N\xi^3\frac{\ln\frac{1}{\zeta^{(t)}}}{t}$, *we have:*

$$\begin{cases}\sum_{t\in[T]}\big(c^{(t)}\big)^2\leq\frac{64N^2\xi^6}{9}\Big[\frac{\pi^2(\ln T_0)^2}{6}+2\Big],\\ \sum_{t\in[T]}\frac{c^{(t)}}{t}\leq\frac{8N\xi^3}{3}\Big[\frac{\pi^2\ln T_0}{6}+1\Big].\end{cases} \quad (B.41)$$

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

$$\begin{cases}\sum_{t\in[T]}(c^{(t)})^2\leq 4N\xi^5\sum_m\mu_m\left[(\ln T_0)^2+(\ln T)^2\right],\\ \sum_{t\in[T]}\frac{c^{(t)}}{t}\leq\sqrt{8N\xi^5\sum_m\mu_m}\Big\{(2+\sqrt{2})\sqrt{\ln T_0}\\ \qquad+\sqrt{2\pi}\,\mathrm{erf}\Big(\sqrt{\frac{\ln T}{2}}\Big)\Big\}.\end{cases} \quad (B.42)$$

Proof of Claim 4: It is easy to check that such $\zeta^{(t)}$ guarantees that $c^{(t)}$ to be decreasing. We can also derive that:

$$\sum_t\delta^{(t)}\leq\frac{2}{T_0}-\frac{1}{T}\leq\frac{2}{T_0}. \quad (B.43)$$

Next we analyze the convergence of $\sum_t(c^{(t)})^2$ and $\sum_t\frac{c^{(t)}}{t}$.

Case 1: $c^{(t)}=\frac{4}{3}N\xi^3\frac{\ln\frac{1}{\zeta^{(t)}}}{t}$.

We start with the case that $c^{(t)}=O\big(\frac{\ln\frac{1}{\delta^{(t)}}}{t}\big)$. As:

$$\sum_{i\geq T_0+1}^{T}\Big(\frac{\ln t}{t}\Big)^2\leq-\frac{(\ln t)^2+2\ln t+2}{t}\Big|_{T_0}^{T}\leq 2, \quad (B.44)$$

$$\sum_{i\in[T_0]}\Big(\frac{\ln T_0}{t}\Big)^2\leq\frac{\pi^2(\ln T_0)^2}{6}, \quad (B.45)$$

we can derive that

$$\sum_{t\in[T]}\big(c^{(t)}\big)^2\leq\frac{64N^2\xi^6}{9}\Big[\sum_{i\in[T_0]}\Big(\frac{\ln T_0}{t}\Big)^2+\sum_{i\geq T_0+1}^{T}\Big(\frac{\ln t}{t}\Big)^2\Big]$$
$$\leq\frac{64N^2\xi^6}{9}\Big[\frac{\pi^2\big(\ln T_0\big)^2}{6}+2\Big]. \quad (B.46)$$

Similarly, as:

$$\sum_{i\geq T_0+1}^{T}\frac{\ln t}{t^2}\leq-\frac{\ln t+1}{t}\Big|_{T_0}^{T}\leq 1, \quad (B.47)$$

$$\sum_{i\in[T_0]}\frac{\ln T_0}{t^2}\leq\frac{\pi^2\ln T_0}{6}, \quad (B.48)$$

we can derive that:

$$\sum_t \frac{c^{(t)}}{t} \leq \frac{8N\xi^3}{3}\left[\frac{\pi^2 \ln T_0}{6}+1\right]. \quad \text{(B.49)}$$

Case 2: $c^{(t)}=2\sqrt{N\xi^5\sum_m \mu_m}\sqrt{\frac{\ln\frac{1}{\delta^{(t)}}}{t}}$. As:

$$\sum_{i\geq T_0+1}^{T} \frac{\ln t}{t} \leq \frac{(\ln t)^2}{2}\Big|_{T_0}^{T} = \frac{1}{2}\left[(\ln T)^2-(\ln T_0)^2\right], \quad \text{(B.50)}$$

$$\sum_{i=1}^{T_0} \frac{\ln T_0}{t} \leq \ln T_0 \ln t\Big|_{1}^{T_0} = (\ln T_0)^2, \quad \text{(B.51)}$$

we can derive that

$$\begin{aligned}\sum_t (c^{(t)})^2 &= 8N\xi^5 \sum_i \mu_i \Big[\sum_{t\in[T_0]} \frac{\ln T_0}{t} + \sum_{t\geq T_0+1}^{T} \frac{\ln t}{t}\Big] \\ &\leq 4N\xi^5 \sum_i \mu_i [(\ln T_0)^2 + (\ln T)^2]. \quad \text{(B.52)}\end{aligned}$$

Similarly, as:

$$\begin{aligned}\sum_{t=T_0+1}^{T} \sqrt{\frac{\ln t}{t^3}} &\leq \sqrt{2\pi}\mathrm{erf}\Big(\sqrt{\frac{\ln t}{2}}\Big) - \sqrt{\frac{\ln t}{t}}\Big|_{T_0}^{T} \\ &\leq \sqrt{2\pi}\mathrm{erf}\Big(\sqrt{\frac{\ln T}{2}}\Big), \quad \text{(B.53)}\end{aligned}$$

$$\begin{aligned}\sum_{t=1}^{T_0} \sqrt{\frac{\ln T_0}{t^3}} &\leq \sum_{t=0}^{\infty} (2^{1-\frac{3}{2}})^t \sqrt{\ln T_0} \\ &\leq (2+\sqrt{2})\ln T_0, \quad \text{(B.54)}\end{aligned}$$

we also have:

$$\begin{aligned}\sum_{t\in[T]} \frac{c^{(t)}}{t} \leq \sqrt{8N\xi^5\sum_i \mu_i}\Big\{&(2+\sqrt{2})\sqrt{\ln T_0} \\ &+\sqrt{2\pi}\mathrm{erf}\Big(\sqrt{\frac{\ln T}{2}}\Big)\Big\}. \quad \text{(B.55)}\end{aligned}$$

Hence, the proof of Claim 4 is done. ∎

Step 5: Now we give the formal regret bound of the SumVar MIS-Learning algorithm.

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 $L_{\mathrm{SumVar}}(\boldsymbol{w})$ in Eq. (10), after T steps of the SumVar algorithm, the choice of $c_n^{(t)}=\epsilon_n^{(t)}$ and*

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

hold the following: when $\frac{\ln\frac{1}{\zeta^{(t)}}}{t} \geq \frac{9\sum_{m\in[N]}\mu_m}{4N\xi}$,

$$\mathbb{E}_{\boldsymbol{x}\sim Q} R_T \leq C_1 \frac{1}{T} + C_2 \frac{\ln T}{T}; \quad \text{(B.56)}$$

otherwise, we have:

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim Q} R_T \leq C_3\frac{1}{T} + C_4\frac{\mathrm{erf}\sqrt{\ln T/2}}{T} + C_5\frac{\ln T}{T} + C_6\frac{(\ln T)^2}{T}. \quad \text{(B.57)}$$

where:

$$\begin{aligned}C_1 = &\Big(\frac{\alpha\eta^2}{2}+\frac{\pi^2\beta^2}{3\alpha\eta^2}\Big) + \Big(\frac{2\xi^2\sum_{i\in[N]}\mu_i\sqrt{N}}{T_0} + \frac{64\beta\xi^3}{3\alpha\eta^2} + \frac{32\pi^2\ln T_0 \beta\xi^3}{9\alpha\eta^2}\Big)N \\ &+ \Big(\frac{256\pi^2(\ln T_0)^2\xi^6}{27\alpha\eta^2} + \frac{1024\xi^6}{9\alpha\eta^2}\Big)N^2, \quad \text{(B.58)}\end{aligned}$$

$$\begin{aligned}C_3 = &\Big(\frac{\alpha\eta^2}{2}+\frac{\pi^2\beta^2}{3\alpha\eta^2}\Big) + \frac{32(\sqrt{2}+1)\beta\sqrt{\ln T_0\xi^5\sum_i\mu_i}}{\alpha\eta^2}\sqrt{N} \\ &+ \left(\frac{2\xi^2\sum_{i\in[N]}\mu_i\sqrt{N}}{T_0} + \frac{32(\ln T_0)^2\xi^5\sum_m\mu_m}{\alpha\eta^2}\right)N, \quad \text{(B.59)}\end{aligned}$$

$$C_2=C_5=\frac{\beta}{2},\ C_4=\frac{32\beta\sqrt{\pi\xi^5\sum_i\mu_i}}{\alpha\eta^2}\sqrt{N},\ C_6=\frac{32\xi^5\sum_i\mu_i}{\alpha\eta^2}N. \quad \text{(B.60)}$$

Proof of Theorem 4: As with probability $N\sum_t \zeta^{(t)} \leq \frac{2N}{T_0}$, we have:

$$\begin{aligned}R_T &= L_{\mathrm{SumVar}}(\boldsymbol{w}) - L_{\mathrm{SumVar}}(\boldsymbol{w}^*) \\ &\leq \nabla L_{\mathrm{SumVar}}(\boldsymbol{w})^\top(\boldsymbol{w}-\boldsymbol{w}^*) \\ &\leq \|\nabla L_{\mathrm{SumVar}}(\boldsymbol{w})\|_2\|\boldsymbol{w}-\boldsymbol{w}^*\|_2 \\ &\leq \sqrt{N}\|\nabla L_{\mathrm{SumVar}}(\boldsymbol{w})\|_\infty, \quad \text{(B.61)}\end{aligned}$$

$$\begin{aligned}\|\nabla L_{\mathrm{SumVar}}(\boldsymbol{w})\|_\infty &= \max_{n\in[N],\boldsymbol{x}\in\Delta}|\mathbb{E}\, Z_n(\boldsymbol{x})| \\ &\leq \xi^2\sum_{i\in[N]}\mu_i. \quad \text{(B.62)}\end{aligned}$$

By plugging bounds of $\sum_t (c^{(t)})^2$ and $\sum_t \frac{c^{(t)}}{t}$ into Claim 3, we can prove Theorem 4. ∎

C. Regret analysis for MinCos MIS-Learning.

We provide the complete analysis for the regret upper bound of MinCos algorithm.

Consider the cost measure $L_{\mathrm{SimCos}}(\boldsymbol{w})$ defined in Eq. (12). Still, let $\boldsymbol{w}^*$ be the optimal mixture:

$$\boldsymbol{w}^* = \mathop{\mathrm{argmin}}_{\boldsymbol{w}\in\Delta} L_{\mathrm{SimCos}}(\boldsymbol{w}) = \mathop{\mathrm{argmin}}_{\boldsymbol{w}\in\Delta}\mathop{\mathrm{argmax}}_{i\in[N]} \ell_i(\boldsymbol{w}), \quad \text{(C.1)}$$

To measure the potential of a moving direction $\boldsymbol{z}$ in decreasing the cost measure $L_{\mathrm{SimCos}}(\boldsymbol{w})$, we introduce a linearization of $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)})$ at $\boldsymbol{w}^{(t)}$:

$$L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}) = \max_{n\in[N]} \ell_n(\boldsymbol{w}^{(t)}) + \nabla\ell_n(\boldsymbol{w}^{(t)})^\top\frac{\boldsymbol{z}-\boldsymbol{w}^{(t)}}{t+1}, \quad \text{(C.2)}$$

which provides an approximation of $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t+1)})$ when $\boldsymbol{w}^{(t+1)}$ is updated by:

$$\boldsymbol{w}^{(t+1)} = \frac{t\boldsymbol{w}^{(t)}+\boldsymbol{z}}{t+1}. \quad \text{(C.3)}$$

We show that this linearization $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z})$ approximate $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t+1)})$ tightly by the following:

Lemma 1. $\left|L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}) - L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t+1)})\right| = O(\frac{\xi^3}{(t+1)^2})$. *More specifically,*

$$L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}) - L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t+1)}) \geq -\frac{\beta'}{2}\left\|\frac{\boldsymbol{z}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2,$$

$$L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}) - L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t+1)}) \leq -\frac{\alpha'}{2}\left\|\frac{\boldsymbol{z}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2.$$

where $\alpha' = \min_{n\in[N]}\alpha_n$, $\beta' = \max_{n\in[N]}\beta_n$.

Proof of Lemma 1: As $L_{\mathrm{SimCos}}(\boldsymbol{w})$ is the pointwise maximum of α_n-strong convexity and β_n smoothness components $\ell_n(\boldsymbol{w}), n\in[N]$, its linearization $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z})$ has the above properties. The proof can be found in (Nesterov, 1998). ∎

Hence, the potential of $\boldsymbol{z}$ in minimizing $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z})$ can approximately measure the potential of $\boldsymbol{z}$ in decreasing $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)})$, and the approximation error decrease at a rate of $\frac{1}{t^2}$. We consider the steepest decrease direction, i.e., the minimizier of Eq. (C.2) as our desired search direction and we claim the search space can be reduced from Δ to $\mathcal{U}$ by the following:

Claim 5.

$\min_{\boldsymbol{z}\in\Delta} L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}) = \min_{\boldsymbol{z}\in\mathcal{U}} L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z})$, *where $\mathcal{U}\triangleq\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_N\}$ represents the standard basis.*

Proof of Claim 5: For $\forall \boldsymbol{y}, \boldsymbol{z} \in \Delta$, let:

$$\mathcal{F}(\boldsymbol{z},\boldsymbol{y}) = \sum_{n\in[N]} y_n \left[\ell_n(\boldsymbol{w}^{(t)}) + \nabla \ell_n(\boldsymbol{w}^{(t)})^\top \frac{\boldsymbol{z}-\boldsymbol{w}^{(t)}}{t+1}\right]. \quad \text{(C.4)}$$

As Δ is convex and compact, and $\mathcal{F}(\boldsymbol{z},\boldsymbol{y})$ is convex on both $\boldsymbol{y}$ and $\boldsymbol{z}$, by the Minima (Sion-Karkutani) Theorem, there exists a saddle point of $\mathcal{F}(\boldsymbol{z},\boldsymbol{y})$ such that:

$$\min_{\boldsymbol{z}\in\Delta}\max_{\boldsymbol{y}\in\Delta} \mathcal{F}(\boldsymbol{z},\boldsymbol{y}) = \max_{\boldsymbol{y}\in\Delta}\min_{\boldsymbol{z}\in\Delta} \mathcal{F}(\boldsymbol{z},\boldsymbol{y}). \quad \text{(C.5)}$$

Let $(\boldsymbol{y}_*, \boldsymbol{z}_*)$ be the such a saddle point. Hence,

$$\boldsymbol{y}_* = \max_{\boldsymbol{y}\in\Delta} \mathcal{F}(\boldsymbol{z}_*,\boldsymbol{y}), \boldsymbol{z}_* = \min_{\boldsymbol{z}\in\Delta} \mathcal{F}(\boldsymbol{z},\boldsymbol{y}_*). \quad \text{(C.6)}$$

As $\mathcal{F}(\boldsymbol{z}_*,\boldsymbol{y})$ is linear on $\boldsymbol{y}$ and $\mathcal{F}(\boldsymbol{z},\boldsymbol{y}_*)$ is linear on $\boldsymbol{z}$, we have $\boldsymbol{y}_*, \boldsymbol{z}_* \in \{\boldsymbol{e}_1, ..., \boldsymbol{e}_N\}$, and:

$$\min_{\boldsymbol{z}\in\Delta}\max_{\boldsymbol{y}\in\Delta} \mathcal{F}(\boldsymbol{z},\boldsymbol{y}) = \min_{\boldsymbol{z}\in\Delta}\max_{n\in[N]} \ell_n(\boldsymbol{w}^{(t)}) + \nabla\ell_n(\boldsymbol{w}^{(t)})^\top \frac{\boldsymbol{z}-\boldsymbol{w}^{(t)}}{t+1}. \quad \text{(C.7)}$$

Hence, we have finished the proof of Claim 5. ∎

Hence, we take $\boldsymbol{e}_{*_t}$, i.e., the steepest decent direction of the linearization $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z})$, as the search direction, and reorganize it as the following:

$$\begin{aligned}\boldsymbol{e}_{*_t} &= \operatorname{argmin}_{\boldsymbol{z}\in\mathcal{U}} L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}) \\ &= \operatorname{argmin}_{\boldsymbol{z}\in\mathcal{U}} L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{z}) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)}).\end{aligned} \quad \text{(C.8)}$$

We estimate $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)})$ as $g_n^{(t)}$ and so estimate $\boldsymbol{e}_{*_t}$ as $\boldsymbol{e}_{I_t}$ where $I_t = \operatorname{argmin}_{n\in[N]} g_n^{(t)}$.

The regret analysis of the SimCos algorithm can be broken down into the following steps.

Step 1: We first partition the regret as:

$$\begin{aligned}&L_{\text{SimCos}}(\boldsymbol{w}^{(t+1)}) - L_{\text{SimCos}}(\boldsymbol{w}^*) \\ \le\, &L_{\text{SimCos}}\left(\frac{t\boldsymbol{w}^{(t)}+\boldsymbol{e}_{I_t}}{t+1}\right) - L_{\text{SimCos}}\left(\frac{t\boldsymbol{w}^{(t)}+\boldsymbol{e}_{*_t}}{t+1}\right) \quad \text{(R1)} \\ +\, &L_{\text{SimCos}}\left(\frac{t\boldsymbol{w}^{(t)}+\boldsymbol{e}_{*_t}}{t+1}\right) - L_{\text{SimCos}}\left(\frac{t\boldsymbol{w}^{(t)}+\boldsymbol{w}^*}{t+1}\right) \quad \text{(R2)} \\ +\, &L_{\text{SimCos}}\left(\frac{t\boldsymbol{w}^{(t)}+\boldsymbol{w}^*}{t+1}\right) - L_{\text{SimCos}}(\boldsymbol{w}^*). \quad \text{(R3)}\end{aligned}$$

The regret can be partitioned into three parts: $R1$ is the *approximation error* and $R2+R3$ is the *optimization error*. In the following, we will bound each part of the regret.

Step 2: We show that $R1$ is upper bounded by:

$$R1 \le 2c_{I_t}^{(t)} + \frac{\beta'}{2}\left\|\frac{\boldsymbol{e}_{I_t}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 - \frac{\alpha'}{2}\left\|\frac{\boldsymbol{e}_{*_t}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2. \quad \text{(C.9)}$$

By Lemma 1, we have:

$$\begin{aligned}R1 \le\, &L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{I_t}) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t}) \\ &+ \frac{\beta'}{2}\left\|\frac{\boldsymbol{e}_{I_t}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 - \frac{\alpha'}{2}\left\|\frac{\boldsymbol{e}_{*_t}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2.\end{aligned} \quad \text{(C.10)}$$

To bound $R1$, we focus on bounding $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{I_t}) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t})$. We first look at $c_n^{(t)}$, i.e., the confidence bound when estimating $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)})$ with $g_n^{(t)}$, which affects the accuracy of estimating $\boldsymbol{e}_{*_t}$ with $\boldsymbol{e}_{I_t}$ when $n=I_t$. The following claim reveals the relationship between $L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{I_t}) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t})$ and $c_{I_t}^{(t)}$:

Claim 6. *Assume $c_n^{(t)}$ satisfies*

$$\mathbb{P}\left[g_n^{(t)} - \left(L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{I_t}) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t})\right) \le c_n^{(t)}\right] \le \zeta^{(t)}, \quad \text{(C.11)}$$

$$\mathbb{P}\left[g_n^{(t)} - \left(L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{I_t}) - L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t})\right) \ge -c_n^{(t)}\right] \le \zeta^{(t)} \quad \text{(C.12)}$$

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

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

Proof of Claim 6: By the union bound, with a probability at least $1-2\zeta^{(t)}$, we have:

$$\begin{aligned}&L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{e}_{I_t}\right) - L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)}\right) \\ &\overset{(a)}{\le} g_{I_t}^{(t)} - c_{I_t}^{(t)} + 2c_{I_t}^{(t)} \overset{(b)}{\le} g_{*_t}^{(t)} - c_{*_t}^{(t)} + 2c_{I_t}^{(t)} \\ &\le L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t}\right) - L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)}\right) + 2c_{I_t}^{(t)}, \end{aligned} \quad \text{(C.13)}$$

where (a) is by the concentration inequality and (b) is by the optimality of $\boldsymbol{e}_{I_t}$. Hence, we finish the proof of Claim 6. ∎

Combining Eq. (C.10) and Claim 6, we finish bounding $R1$.

Step 3: We show that $R2$ is upper bounded by

$$R2 \le -\frac{\alpha'}{2}\left\|\frac{\boldsymbol{w}^*-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 + \frac{\beta'}{2}\left\|\frac{\boldsymbol{e}_{*_t}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 \quad \text{(C.14)}$$

By the optimality of $\boldsymbol{e}_{*_t}$, we can derive that:

$$L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t}\right) \le L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{w}^*\right). \quad \text{(C.15)}$$

Combining with Lemma 1, we have:

$$\begin{aligned}L_{\text{SimCos}}\left(\frac{t\boldsymbol{w}^{(t)}+\boldsymbol{e}_{*_t}}{t+1}\right) &\le L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{e}_{*_t}\right) + \frac{\beta'}{2}\left\|\frac{\boldsymbol{e}_{*_t}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 \\ &\le L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{w}^*\right) + \frac{\beta'}{2}\left\|\frac{\boldsymbol{e}_{*_t}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 \\ &\le L_{\text{SimCos}}\left(\frac{t\boldsymbol{w}^{(t)}+\boldsymbol{w}^*}{t+1}\right) - \frac{\alpha'}{2}\left\|\frac{\boldsymbol{w}^*-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 \\ &\quad + \frac{\beta'}{2}\left\|\frac{\boldsymbol{e}_{*_t}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2. \end{aligned} \quad \text{(C.16)}$$

Hence, we finish bouding $R2$.

Step 4: By the strong convexity and smoothness properties of $\ell_n(\boldsymbol{w})$, we show that $R3$ is upper bounded by:

$$\begin{aligned}R3 \le\, &\frac{t}{t+1}\left[L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)}\right) - L_{\text{SimCos}}(\boldsymbol{w}^*)\right] \\ &+ \frac{\beta'-\alpha'}{2}\left\|\frac{\boldsymbol{w}^*-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2\end{aligned} \quad \text{(C.17)}$$

Let $i = \operatorname{argmax}_{n\in[N]} \ell_n\left(\frac{t\boldsymbol{w}^{(t)}+\boldsymbol{w}^*}{t+1}\right)$, then by the α_i-strongly convexity and β_i-smoothness of $\ell_i(\boldsymbol{w})$:

$$\begin{aligned}&L_{\text{SimCos}}\left(\frac{t\boldsymbol{w}^{(t)}+\boldsymbol{w}^*}{t+1}\right) = \ell_i\left(\frac{t\boldsymbol{w}^{(t)}+\boldsymbol{w}^*}{t+1}\right) \\ &\le \ell_i\left(\boldsymbol{w}^{(t)}\right) + \nabla\ell_i\left(\boldsymbol{w}^{(t)}\right)^\top \frac{\boldsymbol{w}^*-\boldsymbol{w}^{(t)}}{t+1} + \frac{\beta_i}{2}\left\|\frac{\boldsymbol{w}^*-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 \\ &\le \ell_i\left(\boldsymbol{w}^{(t)}\right) + \frac{\ell_i\left(\boldsymbol{w}^*\right) - \ell_i\left(\boldsymbol{w}^{(t)}\right)}{t+1} + \frac{\beta_i-\alpha_i}{2}\left\|\frac{\boldsymbol{w}^*-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 \\ &\le \frac{t}{t+1} L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)}\right) + \frac{1}{t} L_{\text{SimCos}}\left(\boldsymbol{w}^*\right) + \frac{\beta'-\alpha'}{2}\left\|\frac{\boldsymbol{w}^*-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2. \end{aligned} \quad \text{(C.18)}$$

Hence, we finish step 4.

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

$$R_T \leq \frac{2}{T}\sum_{t\in[T-1]} c_{I_t}^{(t)} + 3(\beta'-\alpha')\frac{\ln\frac{T}{2}}{T}. \tag{C.19}$$

Now, combine Eq. (C.9), (C.14) and (C.17), we have:

$$\begin{aligned} R_{t+1} \leq& \frac{t}{t+1}R_t + 2c_{I_t}^{(t)} + \frac{\beta'}{2}\left\|\frac{\boldsymbol{e}_{I_t}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 + \frac{\beta'-\alpha'}{2}\left\|\frac{\boldsymbol{e}_{*_t}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 \\ &+ \frac{2\beta'-\alpha'}{2}\left\|\frac{\boldsymbol{w}^{*}-\boldsymbol{w}^{(t)}}{t+1}\right\|_2^2 \\ \leq& \frac{t}{t+1}R_t + 2c_{I_t}^{(t)} + \frac{3(\beta'-\alpha')}{(t+1)^2}. \end{aligned} \tag{C.20}$$

$$\begin{aligned} TR_T \leq& 2\sum_{t\in[T-1]}\left[c_{I_t}^{(t)} + \frac{3(\beta'-\alpha')}{t+1}\right] \\ \leq& 2\sum_{t\in[T-1]} c_{I_t}^{(t)} + 3(\beta'-\alpha')\ln\frac{T}{2}. \end{aligned} \tag{C.21}$$

Hence, we finish step 5.

Step 6: Next, we focus on bounding $c_n^{(t)}$, which measures the accuracy in estimating $L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n) - L_{\mathrm{SimCos}}(\boldsymbol{w}^{(t)})$ with $g_n^{(t)}$.

In this step, assume $\boldsymbol{x}\sim Q(\boldsymbol{x};\boldsymbol{w}^{(t)})$ for all $\mathbb{E}$ and $\mathbb{V}$ if not specified. To finish this step, we first introduce following components:

$$A_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right) \triangleq \frac{1}{(\mu_i-o_i)^2}\frac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q^2\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right)}, \tag{C.22}$$

$$A_i\left(\boldsymbol{w}^{(t)}\right) \triangleq \frac{1}{(\mu_i-o_i)^2}\sum_{\boldsymbol{x}\in\Omega}\frac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right)} = \mathbb{E}A_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right); \tag{C.23}$$

$$B_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)};n\right) \triangleq \frac{1}{(\mu_i-o_i)^2}\frac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q^2\left(\boldsymbol{w}^{(t)},\boldsymbol{x}\right)}\frac{Q_n(\boldsymbol{x})}{Q\left(\boldsymbol{w}^{(t)},\boldsymbol{x}\right)}, \tag{C.24}$$

$$\begin{aligned} B_i\left(\boldsymbol{w}^{(t)};n\right) \triangleq& \frac{1}{(\mu_i-o_i)^2}\sum_{\boldsymbol{x}\in\Omega}\frac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_n}(\boldsymbol{x})}{Q\left(\boldsymbol{w}^{(t)},\boldsymbol{x}\right)}\frac{Q_n(\boldsymbol{x})}{Q\left(\boldsymbol{w}^{(t)},\boldsymbol{x}\right)}, \\ =& \mathbb{E}B_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)};n\right). \end{aligned} \tag{C.25}$$

By definitions of $\ell_i\left(\boldsymbol{w}^{(t)}\right)$ and $\nabla\ell_i\left(\boldsymbol{w}^{(t)}\right)^{\top}\left(\frac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}\right)$, we have:

$$\ell_i\left(\boldsymbol{w}^{(t)}\right) = A_i\left(\boldsymbol{w}^{(t)}\right) - \frac{\mu_i^2}{(\mu_i-o_i)^2}, \tag{C.26}$$

$$\nabla\ell_i\left(\boldsymbol{w}^{(t)}\right)^{\top}\left(\frac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}\right) = \frac{A_i\left(\boldsymbol{w}^{(t)}\right)-B_i\left(\boldsymbol{w}^{(t)};n\right)}{t+1}. \tag{C.27}$$

Still by the definitions, we can have the following unbiased estimators:

$$\widehat{A}_i\left(\boldsymbol{w}^{(t)}\right) = \frac{1}{(\mu_i-o_i)^2}\frac{1}{t}\sum_{s\in[t]} A_i\left(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)}\right), \tag{C.28}$$

$$\widehat{B}_i\left(\boldsymbol{w}^{(t)};n\right) = \frac{1}{(\mu_i-o_i)^2}\frac{1}{t}\sum_{s\in[t]} B_i\left(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)};n\right), \tag{C.29}$$

$$\widehat{\ell}_i\left(\boldsymbol{w}^{(t)}\right) = \widehat{A}_i\left(\boldsymbol{w}^{(t)}\right) - \frac{\mu_i^2}{(\mu_i-o_i)^2}, \tag{C.30}$$

$$\widehat{\nabla}\ell_i\left(\boldsymbol{w}^{(t)}\right)^{\top}\left(\frac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}\right), = \frac{\widehat{A}_i\left(\boldsymbol{w}^{(t)}\right)-\widehat{B}_i\left(\boldsymbol{w}^{(t)}\right)}{t+1}. \tag{C.31}$$

We first consider concentrations of $\widehat{A}_i\left(\boldsymbol{w}^{(t)}\right)$ and $\widehat{B}_i\left(\boldsymbol{w}^{(t)};n\right)$. For any random variable $X(\boldsymbol{x})$, let $\widetilde{X}(\boldsymbol{x}) \triangleq X(\boldsymbol{x}) - \mathbb{E}X(\boldsymbol{x})$ and define $\varphi(X(\boldsymbol{x})) \triangleq \frac{2\ln(8/\zeta^{(t)})}{3t}\max\widetilde{X}(\boldsymbol{x}) + \sqrt{\frac{2\ln(8/\zeta^{(t)})}{t}\mathbb{V}\widetilde{X}(\boldsymbol{x})}$. We claim that:

Claim C.1. *With the choice of $a_i^{(t)} = \varphi\left(A_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right)\right)$ holds the following:*

$$\mathbb{P}\left[A_i\left(\boldsymbol{w}^{(t)}\right) - \widehat{A}_i\left(\boldsymbol{w}^{(t)}\right) \geq a_i^{(t)}\right] \leq \frac{\zeta^{(t)}}{8}, \tag{C.32}$$

$$\mathbb{P}\left[A_i\left(\boldsymbol{w}^{(t)}\right) - \widehat{A}_i\left(\boldsymbol{w}^{(t)}\right) \leq -a_i^{(t)}\right] \leq \frac{\zeta^{(t)}}{8}. \tag{C.33}$$

$$a_i^{(t)} \leq \frac{2\xi^2}{3(\mu_i-o_i)^2}\frac{\ln\frac{8}{\zeta^{(t)}}}{t} + \frac{\sqrt{2\xi^3\mu_i}}{(\mu_i-o_i)^2}\sqrt{\frac{\ln\frac{8}{\zeta^{(t)}}}{t}}. \tag{C.34}$$

Proof of Claim C.1: Similar to the proof of Theorem 3, we first derive the bounds of $A_i\left(\boldsymbol{w}^{(t)}\right)$, $\mathbb{E}A_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right)$ and $\mathbb{E}A_i^2\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right)$:

$$A_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right) \leq \frac{\xi^2}{(\mu_i-o_i)^2}, \tag{C.35}$$

$$\mathbb{E}A_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right) = A_i\left(\boldsymbol{w}^{(t)}\right) \leq \frac{\xi\mu_i}{(\mu_i-o_i)^2}, \tag{C.36}$$

$$\mathbb{E}A_i^2\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right) \leq \frac{\xi^3\mu_i}{(\mu_i-o_i)^4}. \tag{C.37}$$

By centralizing $A_i\left(\boldsymbol{w}^{(t)}\right)$, we have:

$$\left|\widetilde{A}_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right)\right| \leq \frac{\xi^2}{(\mu_i-o_i)^2}, \tag{C.38}$$

$$\mathbb{E}\widetilde{A}_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right) = 0, \tag{C.39}$$

$$\mathbb{V}\widetilde{A}_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right) \leq \mathbb{E}A_i^2\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right) \leq \frac{\xi^3\mu_i}{(\mu_i-o_i)^4}. \tag{C.40}$$

Then by the Bernstein inequality:

$$\begin{aligned} &\mathbb{P}\left[\widehat{A}_i\left(\boldsymbol{w}^{(t)}\right) - A_i\left(\boldsymbol{w}^{(t)}\right) \geq a_i^{(t)}\right] \\ \leq& e^{-\frac{t\left(a_i^{(t)}\right)^2}{2\mathbb{V}\widetilde{A}_i\left(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)}\right) + \frac{2}{3}a_i^{(t)}\max_{\boldsymbol{x}}\left|\widetilde{A}_i\left(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)}\right)\right|}}, \\ &\mathbb{P}\left[\widehat{A}_i\left(\boldsymbol{w}^{(t)}\right) - A_i\left(\boldsymbol{w}^{(t)}\right) \leq -a_i^{(t)}\right] \\ \leq& e^{-\frac{t\left(a_i^{(t)}\right)^2}{2\mathbb{V}\widetilde{A}_i\left(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)}\right) + \frac{2}{3}a_i^{(t)}\max_{\boldsymbol{x}}\left|\widetilde{A}_i\left(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)}\right)\right|}}. \end{aligned}$$

We finish the proof of Claim C.1 by solving:

$$e^{-\frac{t\left(a_i^{(t)}\right)^2}{2\mathbb{V}\widetilde{A}_i\left(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)}\right) + \frac{2}{3}a_i^{(t)}\max_{\boldsymbol{x}}\left|\widetilde{A}_i\left(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)}\right)\right|}} = \frac{\zeta^{(t)}}{8}, \tag{C.41}$$

and then plugging in the upper bounds of $\left|\widetilde{A}_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right)\right|$ and $\mathbb{V}\widetilde{A}_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right)$. ∎

Claim C.2. *With the choice of $b_{i,n}^{(t)} = \varphi\left(B_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)};n\right)\right)$ holds the following:*

$$\mathbb{P}\left[B_i\left(\boldsymbol{w}^{(t)};n\right) - \widehat{B}_i\left(\boldsymbol{w}^{(t)};n\right) \geq b_{i,n}^{(t)}\right] \leq \frac{\zeta^{(t)}}{8}, \tag{C.42}$$

$$\mathbb{P}\left[B_i\left(\boldsymbol{w}^{(t)};n\right) - \widehat{B}_i\left(\boldsymbol{w}^{(t)};n\right) \leq -b_{i,n}^{(t)}\right] \leq \frac{\zeta^{(t)}}{8}. \tag{C.43}$$

$$b_{i,n}^{(t)} \leq \frac{2\xi^3}{3(\mu_i-o_i)^2}\frac{\ln\frac{8}{\zeta^{(t)}}}{t} + \frac{\sqrt{2\xi^5\mu_i}}{(\mu_i-o_i)^2}\sqrt{\frac{\ln\frac{8}{\zeta^{(t)}}}{t}} \tag{C.44}$$

Proof of Claim C.2: The proof is exactly the same with the proof of Claim C.1. We obtain the bounds of $B_i\left(\boldsymbol{w}^{(t)};n\right)$, $\mathbb{E}B_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)};n\right)$ and $\mathbb{E}B_i^2\left(\boldsymbol{x};\boldsymbol{w}^{(t)};n\right)$:

$$B_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)};n\right) \leq \frac{\xi^3}{(\mu_i-o_i)^2}, \tag{C.45}$$

$$\mathbb{E}B_i\left(\boldsymbol{x};\boldsymbol{w}^{(t)};n\right) = B_i\left(\boldsymbol{w}^{(t)};n\right) \leq \frac{\xi^2\mu_i}{(\mu_i-o_i)^2}, \tag{C.46}$$

$$\mathbb{E}B_i^2\left(\boldsymbol{x};\boldsymbol{w}^{(t)};n\right) \leq \frac{\xi^5\mu_i}{(\mu_i-o_i)^4}. \tag{C.47}$$

By centralizing $B_i\big(\boldsymbol{w}^{(t)};n\big)$, we have:

$$\Big|\widetilde{B}_i\big(\boldsymbol{x};\boldsymbol{w}^{(t)};n\big)\Big|\le\tfrac{\xi^3}{(\mu_i-o_i)^2},\tag{C.48}$$

$$\mathbb{E}\widetilde{B}_i\big(\boldsymbol{x};\boldsymbol{w}^{(t)};n\big)=0,\tag{C.49}$$

$$\mathbb{V}\widetilde{B}_i\big(\boldsymbol{x};\boldsymbol{w}^{(t)};n\big)\le\mathbb{E}B_i^2\big(\boldsymbol{x};\boldsymbol{w}^{(t)};n\big)\le\tfrac{\xi^5\mu_i}{(\mu_i-o_i)^4}.\tag{C.50}$$

The remaining proof follows the same way as the proof of Claim C.1. ∎

Let the active sets of $L_{\text{SimCos}}\big(\boldsymbol{w}^{(t)};\boldsymbol{e}_n\big)$ and $L_{\text{SimCos}}\big(\boldsymbol{w}^{(t)}\big)$ be:

$$I\big(\boldsymbol{w}^{(t)};n\big)=\Big\{k\Big|\ell_k\big(\boldsymbol{w}^{(t)}\big)+\nabla\ell_k\big(\boldsymbol{w}^{(t)}\big)^{\top}\tfrac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}=L_{\text{SimCos}}\big(\boldsymbol{w}^{(t)};\boldsymbol{e}_n\big),k\in[N]\Big\},$$

$$I\big(\boldsymbol{w}^{(t)}\big)=\Big\{k\Big|\ell_k\big(\boldsymbol{w}^{(t)}\big)=L_{\text{SimCos}}\big(\boldsymbol{w}^{(t)}\big),k\in[N]\Big\}.$$

Also, define the complement set by:

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

We then show the convergence rate of $c_n^{(t)}$ by:

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

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

where $A_i(\boldsymbol{x};\boldsymbol{w}^{(t)})=\frac{1}{(\mu_i-o_i)^2}\frac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})}{Q^2\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right)}$,

$$B_i(\boldsymbol{x};\boldsymbol{w}^{(t)};n)=\tfrac{1}{(\mu_i-o_i)^2}\tfrac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_i}(\boldsymbol{x})Q_n(\boldsymbol{x})}{Q^3\left(\boldsymbol{x};\boldsymbol{w}^{(t)}\right)},$$

$$a_i^{(t)}=\varphi\left(A_i(\boldsymbol{x};\boldsymbol{w}^{(t)})\right),b_{i,n}^{(t)}=\varphi\left(B_i(\boldsymbol{x};\boldsymbol{w}^{(t)};n)\right),$$

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

Then, it holds that

$$\mathbb{P}\Big[g_n^{(t)}-(L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n)-L_{\text{SimCos}}(\boldsymbol{w}^{(t)}))\ge\epsilon_n^{(t)}\Big]\le\zeta^{(t)},$$

$$\mathbb{P}\Big[g_n^{(t)}-(L_{\text{SimCos}}(\boldsymbol{w}^{(t)};\boldsymbol{e}_n)-L_{\text{SimCos}}(\boldsymbol{w}^{(t)}))\le-\epsilon_n^{(t)}\Big]\le\zeta^{(t)}.$$

Proof of Theorem 6: For convenience, denote

$$i=\operatorname{argmax}_k\widehat{\ell}_k\big(\boldsymbol{w}^{(t)}\big)+\widehat{\nabla}\ell_k\big(\boldsymbol{w}^{(t)}\big)^{\top}\Big(\tfrac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}\Big);$$

$$j=\operatorname{argmax}_k\widehat{\ell}_k\big(\boldsymbol{w}^{(t)}\big).$$

And let $i'\in I\big(\boldsymbol{w}^{(t)};n\big)$, $j'\in I\big(\boldsymbol{w}^{(t)}\big)$. We have:

$$\begin{aligned}&L_{\text{SimCos}}\big(\boldsymbol{w}^{(t)};\boldsymbol{e}_n\big)-L_{\text{SimCos}}\big(\boldsymbol{w}^{(t)}\big)\\&=\max_{i'}\ell_{i'}\big(\boldsymbol{w}^{(t)}\big)+\nabla\ell_{i'}\big(\boldsymbol{w}^{(t)}\big)^{\top}\Big(\tfrac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}\Big)-\max_{j'}\ell_{j'}\big(\boldsymbol{w}^{(t)}\big)\\&\in\Big[\tfrac{A_{j'}(\boldsymbol{w}^{(t)})-B_{j'}(\boldsymbol{w}^{(t)};n)}{t+1},\tfrac{A_{i'}(\boldsymbol{w}^{(t)})-B_{i'}(\boldsymbol{w}^{(t)};n)}{t+1}\Big];\end{aligned}\tag{C.51}$$

$$\begin{aligned}&\widehat{L}_{\text{SimCos}}\big(\boldsymbol{w}^{(t)};\boldsymbol{e}_n\big)-\widehat{L}_{\text{SimCos}}\big(\boldsymbol{w}^{(t)}\big)\\&=\max_i\widehat{\ell}_i\big(\boldsymbol{w}^{(t)}\big)+\widehat{\nabla}\ell_i\big(\boldsymbol{w}^{(t)}\big)^{\top}\Big(\tfrac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}\Big)-\max_j\widehat{\ell}_j\big(\boldsymbol{w}^{(t)}\big)\\&\in\Big[\tfrac{\widehat{A}_j(\boldsymbol{w}^{(t)})-\widehat{B}_j(\boldsymbol{w}^{(t)};n)}{t+1},\tfrac{\widehat{A}_i(\boldsymbol{w}^{(t)})-\widehat{B}_i(\boldsymbol{w}^{(t)};n)}{t+1}\Big].\end{aligned}\tag{C.52}$$

For $\forall k$, we have:

$$\begin{aligned}&\Big|A_k\Big(\boldsymbol{w}^{(t)}\Big)-B_k\Big(\boldsymbol{w}^{(t)};n\Big)\Big|=\Big|\nabla\ell_k\Big(\boldsymbol{w}^{(t)}\Big)^{\top}\big(\boldsymbol{e}_n-\boldsymbol{w}^{(t)}\big)\Big|\\&=\tfrac{1}{(\mu_k-o_k)^2}\Big|-\textstyle\sum_{\boldsymbol{x}\in\Omega}\tfrac{P^2(\boldsymbol{x})\mathbf{1}_{\mathcal{E}_k}(\boldsymbol{x})}{Q^2(\boldsymbol{x};\boldsymbol{w}^{(t)})}\boldsymbol{Q}(\boldsymbol{x})^{\top}\big(\boldsymbol{e}_n-\boldsymbol{w}^{(t)}\big)\Big|\\&\le\tfrac{1}{(\mu_k-o_k)^2}\xi^2\mu_k\left\|\boldsymbol{e}_n-\boldsymbol{w}^{(t)}\right\|_1\le\tfrac{2\xi^2\mu_k}{(\mu_k-o_k)^2},\end{aligned}\tag{C.53}$$

$$\begin{aligned}&\Big|\widehat{A}_k\Big(\boldsymbol{w}^{(t)}\Big)-\widehat{B}_k\Big(\boldsymbol{w}^{(t)};n\Big)\Big|=\Big|\widehat{\nabla}\ell_k\Big(\boldsymbol{w}^{(t)}\Big)^{\top}\big(\boldsymbol{e}_n-\boldsymbol{w}^{(t)}\big)\Big|\\&=\tfrac{1}{(\mu_k-o_k)^2}\Big|-\tfrac{1}{t}\textstyle\sum_{s\in[t]}\tfrac{P^2(\boldsymbol{x}^{(s)})\mathbf{1}_{\mathcal{E}_k}(\boldsymbol{x}^{(s)})}{Q^3(\boldsymbol{x}^{(s)};\boldsymbol{w}^{(t)})}\boldsymbol{Q}(\boldsymbol{x}^{(s)})^{\top}\big(\boldsymbol{e}_n-\boldsymbol{w}^{(t)}\big)\Big|\\&\le\tfrac{1}{(\mu_k-o_k)^2}\xi^3\left\|\boldsymbol{e}_n-\boldsymbol{w}^{(t)}\right\|_1\le\tfrac{2\xi^3}{(\mu_k-o_k)^2}.\end{aligned}\tag{C.54}$$

Therefore, combine Eq. (C.51), (C.52), (C.53) and (C.54), we always have:

$$\begin{aligned}&\Big|\Big\{L_{\text{SimCos}}\big(\boldsymbol{w}^{(t)};\boldsymbol{e}_n\big)-L_{\text{SimCos}}\big(\boldsymbol{w}^{(t)}\big)\Big\}-g_n^{(t)}\Big|\\&\le\tfrac{1}{t+1}\Big[\max_{k\in[N]}\tfrac{2\xi^2\mu_k}{(\mu_k-o_k)^2}+\max_{k\in[N]}\tfrac{2\xi^3}{(\mu_k-o_k)^2}\Big].\end{aligned}\tag{C.55}$$

Hence, $\epsilon_n^{(t)}\le\frac{C_1}{t+1}$ always hold. We have finished the proof for the case that $\epsilon_n^{(t)}=\frac{C_1}{t+1}$.

Next, we analyze when $\epsilon_n^{(t)}$ has a faster convergence rate. By the definition of ξ-similarity, we show that $\forall k$, the following relationship between $B_k\big(\boldsymbol{w}^{(t)};n\big)$ and $A_k\big(\boldsymbol{w}^{(t)}\big)$ exists:

$$B_k\big(\boldsymbol{w}^{(t)};n\big)\in\Big[\max\{\tfrac{1}{\xi},w_n^{(t)}\}A_k\big(\boldsymbol{w}^{(t)}\big),\xi A_k\big(\boldsymbol{w}^{(t)}\big)\Big].\tag{C.56}$$

Similarly, we have:

$$\widehat{B}_k\big(\boldsymbol{w}^{(t)};n\big)\in\Big[\max\{\tfrac{1}{\xi},w_h^{(t)}\}\widehat{A}_k\big(\boldsymbol{w}^{(t)}\big),\xi\widehat{A}_k\big(\boldsymbol{w}^{(t)}\big)\Big].\tag{C.57}$$

Also, due to the optimality of i and j, we have:

$$\widehat{\ell}_i\big(\boldsymbol{w}^{(t)}\big)\le\widehat{\ell}_j\big(\boldsymbol{w}^{(t)}\big),\tag{C.58}$$

$$\begin{aligned}&\widehat{\ell}_i\big(\boldsymbol{w}^{(t)}\big)+\widehat{\nabla}\ell_i\big(\boldsymbol{w}^{(t)}\big)^{\top}\Big(\tfrac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}\Big)\\&\ge\widehat{\ell}_j\big(\boldsymbol{w}^{(t)}\big)+\widehat{\nabla}\ell_j\big(\boldsymbol{w}^{(t)}\big)^{\top}\Big(\tfrac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}\Big),\end{aligned}\tag{C.59}$$

which imply that:

$$\widehat{A}_i\big(\boldsymbol{w}^{(t)}\big)-\tfrac{\mu_i^2}{(\mu_i-o_i)^2}\le\widehat{A}_j\big(\boldsymbol{w}^{(t)}\big)-\tfrac{\mu_j^2}{(\mu_j-o_j)^2},\tag{C.60}$$

$$\begin{aligned}&\widehat{A}_i\big(\boldsymbol{w}^{(t)}\big)-\tfrac{\mu_i^2}{(\mu_i-o_i)^2}+\tfrac{\widehat{A}_i(\boldsymbol{w}^{(t)})-\widehat{B}_i(\boldsymbol{w}^{(t)};n)}{t+1}\\&\ge\widehat{A}_j\big(\boldsymbol{w}^{(t)}\big)-\tfrac{\mu_j^2}{(\mu_j-o_j)^2}+\tfrac{\widehat{A}_j(\boldsymbol{w}^{(t)})-\widehat{B}_j(\boldsymbol{w}^{(t)};n)}{t+1}.\end{aligned}\tag{C.61}$$

Combine Eq. (C.57) and (C.61), we have:

$$\begin{aligned}&\tfrac{t+2-\max\left\{\frac{1}{\xi},w_n^{(t)}\right\}}{t+1}\widehat{A}_i\big(\boldsymbol{w}^{(t)}\big)-\tfrac{\mu_i^2}{(\mu_i-o_i)^2}\\&\ge\widehat{A}_i\big(\boldsymbol{w}^{(t)}\big)-\tfrac{\mu_i^2}{(\mu_i-o_i)^2}+\tfrac{\widehat{A}_i(\boldsymbol{w}^{(t)})-\widehat{B}_i(\boldsymbol{w}^{(t)};h)}{t+1}\\&\ge\widehat{A}_j\big(\boldsymbol{w}^{(t)}\big)-\tfrac{\mu_j^2}{(\mu_j-o_j)^2}+\tfrac{\widehat{A}_j(\boldsymbol{w}^{(t)})-\widehat{B}_j(\boldsymbol{w}^{(t)};n)}{t+1}\\&\ge\tfrac{t+2-\xi}{t+1}\widehat{A}_j\big(\boldsymbol{w}^{(t)}\big)-\tfrac{\mu_j^2}{(\mu_j-o_j)^2}.\end{aligned}\tag{C.62}$$

Combine Eq. (C.60) and (C.62), we have:

$$\tfrac{t+2-\xi}{t+2-\max\left\{\frac{1}{\xi},w_n^{(t)}\right\}}\widehat{A}_j\big(\boldsymbol{w}^{(t)}\big)$$

$$+\frac{t+1}{t+2-\max\left\{\frac{1}{\xi},w_n^{(t)}\right\}}\left[\frac{\mu_i^2}{(\mu_i-o_i)^2}-\frac{\mu_j^2}{(\mu_j-o_j)^2}\right]$$
$$\leq\widehat{A}_i\left(\boldsymbol{w}^{(t)}\right)\leq\widehat{A}_j\left(\boldsymbol{w}^{(t)}\right)+\left[\frac{\mu_i^2}{(\mu_i-o_i)^2}-\frac{\mu_j^2}{(\mu_j-o_j)^2}\right]. \quad \text{(C.63)}$$

Eq. (C.63) implies that $i{=}j$ when t is large enough such that:

$$t\geq\max_{i,j}\frac{(\xi-\frac{1}{\xi})\widehat{A}_j\left(\boldsymbol{w}^{(t)}\right)+(1-\frac{1}{\xi})\left[\frac{\mu_i^2}{(\mu_i-o_i)^2}-\frac{\mu_j^2}{(\mu_j-o_j)^2}\right]}{\widehat{A}_i\left(\boldsymbol{w}^{(t)}\right)-\frac{\mu_i^2}{(\mu_i-o_i)^2}-\widehat{A}_j\left(\boldsymbol{w}^{(t)}\right)+\frac{\mu_j^2}{(\mu_j-o_j)^2}}-2+\frac{1}{\xi}. \quad \text{(C.64)}$$

We can relax the right hand side of Eq. (C.64) and let:

$$t\geq\max_{i\neq j}\frac{(\xi-\frac{1}{\xi})\left[A_j\left(\boldsymbol{w}^{(t)}\right)+a_j^{(t)}\right]+(1-\frac{1}{\xi})\left[\frac{\mu_i^2}{(\mu_i-o_i)^2}-\frac{\mu_j^2}{(\mu_j-o_j)^2}\right]}{\left[A_i\left(\boldsymbol{w}^{(t)}\right)-a_i^{(t)}-\frac{\mu_i^2}{(\mu_i-o_i)^2}\right]-\left[A_j\left(\boldsymbol{w}^{(t)}\right)+a_j^{(t)}-\frac{\mu_j^2}{(\mu_j-o_j)^2}\right]}$$
$$-2+\frac{1}{\xi}. \quad \text{(C.65)}$$

By Claim C.1, we can show that Eq. (C.64) is satisfied with a probability at least $1-\frac{\zeta^{(t)}}{4}$ if Eq. (C.65) is satisfied.

Combine Claim C.1, C.2, and Eq. (C.26), (C.27), we see that when t is large enough, the estimation for $A_k\left(\boldsymbol{w}^{(t)}\right)$, $B_k\left(\boldsymbol{w}^{(t)};n\right)$, $\ell_k\left(\boldsymbol{w}^{(t)}\right)$, and $\nabla\ell_k\left(\boldsymbol{w}^{(t)}\right)^\top\left(\frac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}\right)$ will be accurate enough, and so we will be able to distinguish $I\left(\boldsymbol{w}^{(t)};n\right)$ from $I^c\left(\boldsymbol{w}^{(t)};n\right)$, and $I\left(\boldsymbol{w}^{(t)}\right)$ from $I^c\left(\boldsymbol{w}^{(t)}\right)$ with a high probability.

Namely, $i{\in}I\left(\boldsymbol{w}^{(t)};h\right)$ with a probability at least $1-\frac{\zeta^{(t)}}{2}$, when t is large enough such that

$$\frac{2(t+2)a_i^{(t)}}{t+1}+\frac{2b_{i,n}^{(t)}}{t+1}\leq L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{e}_n\right)$$
$$-\max_{k\in I^c\left(\boldsymbol{w}^{(t)};n\right)}\left[\frac{t+2}{t+1}A_k\left(\boldsymbol{w}^{(t)}\right)-B_k\left(\boldsymbol{w}^{(t)};n\right)-\frac{\mu_k^2}{(\mu_k-o_k)^2}\right]. \quad \text{(C.66)}$$

And, $j{\in}I\left(\boldsymbol{w}^{(t)}\right)$ with a probability at least $1-\frac{\zeta^{(t)}}{4}$, when t is large enough such that

$$2a_j^{(t)}\leq L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)}\right)$$
$$-\max_{k\in I^c\left(\boldsymbol{w}^{(t)}\right)}\left[A_k\left(\boldsymbol{w}^{(t)}\right)-\frac{\mu_k^2}{(\mu_k-o_k)^2}\right]. \quad \text{(C.67)}$$

To summarize, if t satisfies Eq.(C.65), (C.66) and (C.67), then with a probability at least $1-\zeta^{(t)}$ that

$$\left\{L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{e}_n\right)-L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)}\right)\right\}$$
$$-\left\{\widehat{L}_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{e}_h\right)-\widehat{L}_{\text{SimCos}}\left(\boldsymbol{w}^{(t)}\right)\right\}$$
$$=\left\{\ell_{i'}\left(\boldsymbol{w}^{(t)}\right)+\nabla\ell_{i'}\left(\boldsymbol{w}^{(t)}\right)^\top\frac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}-\ell_{j'}\left(\boldsymbol{w}^{(t)}\right)\right\}$$
$$-\left\{\hat{\ell}_i\left(\boldsymbol{w}^{(t)}\right)+\hat{\nabla}\ell_i\left(\boldsymbol{w}^{(t)}\right)^\top\frac{\boldsymbol{e}_n-\boldsymbol{w}^{(t)}}{t+1}-\hat{\ell}_j\left(\boldsymbol{w}^{(t)}\right)\right\}$$
$$\overset{(a)}{=}\left\{\ell_i\left(\boldsymbol{w}^{(t)}\right)+\nabla\ell_i\left(\boldsymbol{w}^{(t)}\right)^\top\frac{\boldsymbol{e}_h-\boldsymbol{w}^{(t)}}{t+1}-\ell_j\left(\boldsymbol{w}^{(t)}\right)\right\}$$
$$-\left\{\hat{\ell}_i\left(\boldsymbol{w}^{(t)}\right)+\hat{\nabla}\ell_i\left(\boldsymbol{w}^{(t)}\right)^\top\frac{\boldsymbol{e}_h-\boldsymbol{w}^{(t)}}{t+1}-\hat{\ell}_j\left(\boldsymbol{w}^{(t)}\right)\right\}$$
$$\overset{(b)}{=}\left\{\ell_i\left(\boldsymbol{w}^{(t)}\right)+\nabla\ell_i\left(\boldsymbol{w}^{(t)}\right)^\top\frac{\boldsymbol{e}_h-\boldsymbol{w}^{(t)}}{t+1}-\ell_i\left(\boldsymbol{w}^{(t)}\right)\right\}$$
$$-\left\{\hat{\ell}_i\left(\boldsymbol{w}^{(t)}\right)+\hat{\nabla}\ell_i\left(\boldsymbol{w}^{(t)}\right)^\top\frac{\boldsymbol{e}_h-\boldsymbol{w}^{(t)}}{t+1}-\hat{\ell}_i\left(\boldsymbol{w}^{(t)}\right)\right\}$$
$$\overset{(c)}{=}\frac{A_i\left(\boldsymbol{w}^{(t)}\right)-B_i\left(\boldsymbol{w}^{(t)};n\right)}{t+1}-\frac{\widehat{A}_i\left(\boldsymbol{w}^{(t)}\right)-\widehat{B}_i\left(\boldsymbol{w}^{(t)};n\right)}{t+1}$$
$$\leq\frac{a_i^{(t)}+b_{i,h}^{(t)}}{t+1}. \quad \text{(C.68)}$$

Namely, if t satisfies Eq. (C.65), (C.66) and (C.67), then with a probability at least $1-\zeta^{(t)}$ that

$$tL_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{e}_n\right)-L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)}\right)-g_n^{(t)}\leq\frac{a_i^{(t)}+b_{i,n}^{(t)}}{t+1}. \quad \text{(C.69)}$$

We assume $t{\geq}T_0$ is large enough to satisfy above conditions and we leave the discussion on the selection of T_0 to the proof of Theorem 7.

Note that when Eq. (C.65), (C.66) and (C.67) are satisfied, (a) is achieved because that with probabilities at least $1-\frac{\zeta^{(t)}}{2}$ and $1-\frac{\zeta^{(t)}}{4}$, we have $i{\in}I\left(\boldsymbol{w}^{(t)};n\right)$ and $j{\in}I\left(\boldsymbol{w}^{(t)}\right)$; (b) is achieved for with a probability at least $1-\frac{\zeta^{(t)}}{4}$, we have $i{=}j$; (c) can be proved by plugging $A_i\left(\boldsymbol{w}^{(t)}\right)$, $B_i\left(\boldsymbol{w}^{(t)};n\right)$, $\widehat{A}_k\left(\boldsymbol{w}^{(t)}\right)$ and $\widehat{B}_k\left(\boldsymbol{w}^{(t)};n\right)$.

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

$$L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{e}_n\right)-L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)}\right)-g_n^{(t)}$$
$$=\frac{A_i\left(\boldsymbol{w}^{(t)}\right)-B_i\left(\boldsymbol{w}^{(t)};h\right)}{t+1}-\frac{\widehat{A}_i\left(\boldsymbol{w}^{(t)}\right)-\widehat{B}_i\left(\boldsymbol{w}^{(t)};n\right)}{t+1}$$
$$\geq-\frac{a_i^{(t)}+b_{i,n}^{(t)}}{t+1}. \quad \text{(C.70)}$$

Namely, when t is large enough to satisfy Eq. (C.65), (C.66) and (C.67), $\epsilon_n^{(t)}\leq\max_{i\in[N]}\frac{a_i^{(t)}+b_{i,n}^{(t)}}{t+1}$. ∎

By Theorem 6, we have the upper bounds of $c_n^{(t)}$ under different conditions which depend on C_1, $a_i^{(t)}$ and $b_{i,n}^{(t)}$. The bound of C_1 is given by Theorem 6. and the bounds of $a_i^{(t)}$ and $b_{i,n}^{(t)}$ are given by Claim C.1 and C.2.

Step 7: Now we give the formal regret bound of the SimCos MIS-Learning algorithm.

Assume that:

$$\rho=\min_{\boldsymbol{w}^{(t)},h}L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)};\boldsymbol{e}_h\right)$$
$$-\max_{k\in I^c\left(\boldsymbol{w}^{(t)};h\right)}\ell_k\left(\boldsymbol{w}^{(t)};\boldsymbol{e}_h\right)+\nabla\ell_k\left(\boldsymbol{w}^{(t)}\right)^\top\frac{\boldsymbol{e}_h-\boldsymbol{w}^{(t)}}{t+1}, \quad \text{(C.71)}$$

$$\gamma=\min_{\boldsymbol{w}^{(t)}}L_{\text{SimCos}}\left(\boldsymbol{w}^{(t)}\right)-\max_{k\in I^c\left(\boldsymbol{w}^{(t)}\right)}\ell_k\left(\boldsymbol{w}^{(t)}\right). \quad \text{(C.72)}$$

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_{SimCos} in Eq. (12), after T steps of the SimCos algorithm, the choice of $c_n^{(t)}{=}\epsilon_n^{(t)}$ and*

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

hold the following:

$$\mathbb{E}R_T\leq2\left[C_1T_0-C_2(\ln\sqrt{8}T_0)^2-2C_3\sqrt{T_0\ln8T_0^2}\right]\cdot\frac{1}{T}$$

$$+3(\beta'-\alpha')\tfrac{\ln T}{T}+2C_2\tfrac{(\ln\sqrt{8}T)^2}{T}+4C_3\sqrt{\tfrac{\ln 8T^2}{T}}, \tag{C.73}$$

where

$$C_1=\max_{k\in[N]}\tfrac{2\xi^2\mu_k}{(\mu_k-o_k)^2}+\max_{k\in[N]}\tfrac{2\xi^3}{(\mu_k-o_k)^2}, \tag{C.74}$$

$$C_2=\max_{k\in[N]}\tfrac{2\xi^2(1+\xi)}{3(\mu_k-o_k)^2}, \tag{C.75}$$

$$C_3=\max_{k\in[N]}\tfrac{\sqrt{2\xi^3\mu_i}(1+\xi)}{(\mu_i-o_i)^2}, \tag{C.76}$$

$$C_4=\max\nolimits_{i,j\in[N]}\tfrac{(\frac{2}{\xi}+1)(\xi^2-1)\mu_j}{\gamma(\mu_j-o_j)^2}+\tfrac{(1-\frac{1}{\xi})(1+\frac{2}{\xi})\mu_i^2}{\gamma(\mu_i-o_i)^2}, \tag{C.77}$$

$$C_5=\min\left\{\tfrac{\rho(C_4+1)(\xi+1)}{2(C_4+\xi+2)},\tfrac{\gamma}{2+\xi}\right\}, \tag{C.78}$$

$$T_0=\max\left\{C_4,\left(\tfrac{C_3+\sqrt{C_3^2+4C_3C_5}}{2C_5}\right)^4,150\right\} \tag{C.79}$$

Proof of Theorem 7: Let T_0-1 be the last time before T such that Eq. (C.65), (C.66) and (C.67) are not all satisfied. Then,

$$\begin{aligned}
&\mathbb{E}\textstyle\sum_{t\in[T]}c_{I_t}^{(t)}\\
&\overset{(a)}{\le}C_1(T_0-1+\tfrac{2}{T_0})+\textstyle\sum_{t=T_0}^{T}\max_{k\in[N]}a_k^{(t)}+b_{k,I_t}^{(t)}\\
&\overset{(b)}{\le}C_1T_0+C_2\textstyle\sum_{t=T_0}^{T}\tfrac{\ln\frac{8}{\zeta^{(t)}}}{t}+C_3\textstyle\sum_{t=T_0}^{T}\sqrt{\tfrac{\ln\frac{8}{\zeta^{(t)}}}{t}},\\
&\overset{(c)}{\le}C_1T_0+C_2\left.\left(\ln\sqrt{8}t\right)^2\right|_{T_0}^{T}+2C_3\left.\sqrt{t\ln 8t^2}\right|_{T_0}^{T},\\
&=\left[C_1T_0-C_2(\ln\sqrt{8}T_0)^2-2C_3\sqrt{T_0\ln 8T_0^2}\right]\\
&\quad+C_2(\ln\sqrt{8}T)^2+2C_3\sqrt{T\ln 8T^2}.
\end{aligned} \tag{C.80}$$

Notice that: (a) is achieved by Theorem 6 and $\sum_t\zeta^{(t)}\le\frac{2}{T_0}$; (b) is achieved by plugging the upper bounds of $a_k^{(t)}$ and $b_{k,I_t}^{(t)}$ in Claim C.1 and C.2; and (c) is achieved for:

$$\sum_{t=a}^{b}\frac{\ln\frac{8}{\zeta^{(t)}}}{t}<\left.\left(\ln\sqrt{8}t\right)^2\right|_a^b, \tag{C.81}$$

$$\begin{aligned}
\sum_{t=a}^{b}\sqrt{\frac{\ln\frac{8}{\zeta^{(t)}}}{t}}&<\left.2\sqrt{t\ln 8t^2}-4\sqrt{t}D_+\left(\tfrac{1}{2}\sqrt{\ln 8t^2}\right)\right|_a^b\\
&<\left.2\sqrt{t\ln 8t^2}\right|_a^b.
\end{aligned} \tag{C.82}$$

where $D_+(\cdot)$ is the Dawson's integral.

Combine Eq. (C.21) and (C.80), we have Eq. (C.73). Next, we analyze the upper bound of T_0. By the upper bounds $a_k^{(t)},b_{k,h}^{(t)}$ given by Claim C.1 and Claim C.2 and Eq. (C.75) and (C.76), we have $\forall k\in[N]$ that:

$$\tfrac{2(t+2)a_k^{(t)}}{t+1}+\tfrac{2b_{k,h}^{(t)}}{t+1}\le\tfrac{2(t+\xi+2)}{(t+1)(\xi+1)}\left(C_2\tfrac{\ln\frac{8}{\zeta^{(t)}}}{t}+C_3\sqrt{\tfrac{\ln\frac{8}{\zeta^{(t)}}}{t}}\right). \tag{C.83}$$

Hence, it is sufficient to say Eq. (C.66) is satisfied if:

$$\tfrac{2(t+\xi+2)}{(t+1)(\xi+1)}\left(C_2\tfrac{\ln\frac{8}{\zeta^{(t)}}}{t}+C_3\sqrt{\tfrac{\ln\frac{8}{\zeta^{(t)}}}{t}}\right)\le\rho. \tag{C.84}$$

Similarly, it is sufficient to say Eq. (C.67) is satisfied if:

$$2\left(C_2\tfrac{\ln\frac{8}{\zeta^{(t)}}}{t}+C_3\sqrt{\tfrac{\ln\frac{8}{\zeta^{(t)}}}{t}}\right)\le\gamma. \tag{C.85}$$

With Eq. (C.72), in the right hand side of Eq. (C.65),

$$\begin{aligned}
&\left[A_i\left(\boldsymbol{w}^{(t)}\right)-\tfrac{\mu_i^2}{(\mu_i-o_i)^2}\right]-\left[A_j\left(\boldsymbol{w}^{(t)}\right)-\tfrac{\mu_j^2}{(\mu_j-o_j)^2}\right]\\
=&L_{\mathrm{SimCos}}\left(\boldsymbol{w}^{(t)}\right)-\ell_j\left(\boldsymbol{w}^{(t)}\right)\ge\gamma.
\end{aligned} \tag{C.86}$$

Notice that $A_j\left(\boldsymbol{w}^{(t)}\right)\le\frac{\xi\mu_j}{(\mu_j-o_j)^2}$. By relaxing Eq. (C.85) to:

$$(2+\xi)\left(C_2\tfrac{\ln\frac{8}{\zeta^{(t)}}}{t}+C_3\sqrt{\tfrac{\ln\frac{8}{\zeta^{(t)}}}{t}}\right)\le\gamma, \tag{C.87}$$

we have $\forall k$, $(2+\xi)a_k^{(t)}\le\gamma$. Hence, the right hand side of Eq. (C.65) is upper bounded by:

$$\begin{aligned}
&\frac{\frac{(\xi^2-1)\mu_j-(1-\frac{1}{\xi})\mu_j^2}{(\mu_j-o_j)^2}+\frac{(1-\frac{1}{\xi})\mu_i^2}{(\mu_i-o_i)^2}+(\xi-\frac{1}{\xi})a_j^{(t)}}{\gamma-a_i^{(t)}-a_j^{(t)}}-2+\frac{1}{\xi}\\
\le&\frac{(\frac{2}{\xi}+1)(\xi^2-1)\mu_j}{\gamma(\mu_j-o_j)^2}+\frac{(1-\frac{1}{\xi})(1+\frac{2}{\xi})\mu_i^2}{\gamma(\mu_i-o_i)^2}\\
\le&\max_{i,j\in[N]}\frac{(\frac{2}{\xi}+1)(\xi^2-1)\mu_j}{\gamma(\mu_j-o_j)^2}+\frac{(1-\frac{1}{\xi})(1+\frac{2}{\xi})\mu_i^2}{\gamma(\mu_i-o_i)^2}\triangleq C_4.
\end{aligned} \tag{C.88}$$

Hence, it is sufficient to say Eq. (C.65) is achieved if $t\ge C_4$. As $\forall t\ge150$, $\frac{\ln 8t^2}{t}\le\frac{1}{\sqrt{t}}$. By solving $C_2t^{-\frac{1}{2}}+C_3t^{-\frac{1}{4}}=C_5$:

$$t=\left(\frac{C_3+\sqrt{C_3^2+4C_3C_5}}{2C_5}\right)^4, \tag{C.89}$$

which guarantees both Eq. (C.84) and (C.85) can be satisfied. Hence, T_0 is upper bounded by:

$$T_0\le\max\left\{C_4,\left(\frac{C_3+\sqrt{C_3^2+4C_3C_5}}{2C_5}\right)^4,150\right\}. \tag{C.90}$$

■