On Approximating Total Variation Distance (Paper #541)

Arnab Bhattacharyya[1], Sutanu Gayen[2], Kuldeep S. Meel[1], Dimitrios Myrisiotis[1], A. Pavan[3], and N. V. Vinodchandran[4]

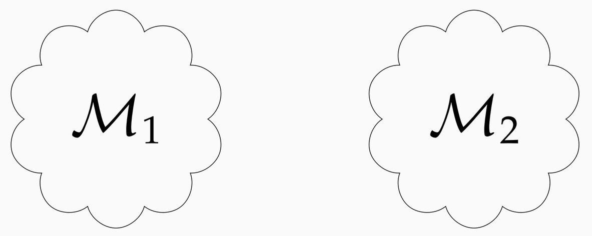
[1]National University of Singapore; [2]Indian Institute of Technology Kanpur; [3]Iowa State University; [4]University of Nebraska-Lincoln

Introduction

The study of probabilistic models lies at the heart of statistical machine learning. Examples of probabilistic models:

1. Probabilistic graphical models (Ising models, Bayes nets, etc.);
2. Deep generative models (autoencoders, GANs, etc.).

Problem (Distance computation problem; informal). *Given two probabilistic models $\mathcal{M}_1, \mathcal{M}_2$, compute/estimate how close are they.*



There are many notions of distance between distributions:

1. f-divergences (Hellinger, KL, χ^2, etc.);
2. Integral probability metrics (Wasserstein, TV, etc.).

We focus on *total variation (TV) distance* d_{TV}.

Definition. *For distributions P, Q over a common domain D,*

$$d_{\mathrm{TV}}(P,Q) := \sup_{A \subseteq D} |P(A) - Q(A)|\,.$$

TV distance is important, because:

1. It is natural: $d_{\mathrm{TV}}(P,Q)$ is equal to the maximum gap between the probabilities assigned by P and Q to a single event;
2. It has many desirable properties: It is a metric, it is bounded in $[0,1]$, and is invariant with respect to bijections.

Central Question

What is the computational complexity of computing/estimating TV distance between two probabilistic models?

Results

Product distributions are distributions of the form $P = \bigotimes_{i=1}^{n} \mathrm{Bern}(p_i)$ where $p_1, \ldots, p_n \in [0,1]$.

Theorem 1. *Computing the TV distance between two product distributions is #P-hard.*

Theorem 2. *There is an FPTAS for estimating the TV distance between an arbitrary product distribution P and a product distribution Q with a bounded number of distinct parameters.*

Related Work

- Goldreich, Sahai, and Vadhan (1999, 2003) showed that TV distance is hard to additively estimate for distributions samplable by Boolean circuits.
- Canonne and Rubinfeld (2014) showed how to additively estimate TV distance for probabilistic models with efficient inference and sampling.
- Feng, Guo, Jerrum, and Wang (2022) designed an FPRAS for estimating the TV distance between two product distributions. (This work is a *follow-up* to our work.)

Techniques: Theorem 1

Definition (#SubsetProd). *Given integers $a_1, \ldots, a_n$, and a target integer T, compute the number of sets $S \subseteq [n]$ such that $\prod_{i \in S} a_i = T$.*

Theorem (Garey and Johnson (1979); Yao). *#SubsetProd is #P-hard.*

Definition (#PMFEquals). *Given $p_1, \ldots, p_n \in [0,1]$ and a target $v \in [0,1]$, compute the number of $x \in \{0,1\}^n$ such that $P(x) = v$, where P is the product distribution defined by $p_1, \ldots, p_n$.*

Claim. #SubsetProd $\overset{(1)}{\leq_m}$ #PMFEquals $\overset{(2)}{\leq_T}$ d_{TV}.

(1): **#SubsetProd $\leq_m$ #PMFEquals.** Let $p_i := a_i / (1 + a_i)$ and $v := T \prod_{i=1}^{n} (1 - p_i)$. Then

$$\prod_{i \in S} a_i = T \Leftrightarrow \prod_{i \in S} \frac{p_i}{1 - p_i} = \frac{v}{\prod_{i \in [n]} (1 - p_i)} \Leftrightarrow P(1_S) = v.$$

(2): **#PMFEquals $\leq_T$ d_{TV}.** Define auxiliary distributions $P', Q', \hat{P}, \hat{Q}$ as follows:

- $\hat{p}_i := p_i$ for $i \in [n]$ and $\hat{p}_{n+1} := 1$; $\hat{q}_i := 1/2$ for $i \in [n]$ and $\hat{q}_{n+1} := v2^n$;
- $p'_i := p_i$ for $i \in [n]$, $p'_{n+1} := 1$, and $p'_{n+2} := \frac{1}{2} + \beta$; $q'_i := \frac{1}{2}$ for $i \in [n]$, $q'_{n+1} := v2^n$, and $q'_{n+2} := \frac{1}{2} - \beta$ for some appropriately chosen β that depends on the granularity of our precision.

Claim. *It is the case that*

$$|\{x \mid P(x) = v\}| = \left(d_{\mathrm{TV}}(P', Q') - d_{\mathrm{TV}}(\hat{P}, \hat{Q})\right) / (2\beta v)\,.$$

Techniques: Theorem 2

For simplicity of presentation, consider the case where Q is the uniform distribution $\mathbb{U}$. The idea is to reduce the computation of $d_{\mathrm{TV}}(P, \mathbb{U})$ to $O(\mathrm{poly}(n))$ instances of #Knapsack. Since the latter problem has an FPTAS by Gopalan, Klivans, and Meka (2010), and Stefankovic, Vempala, and Vigoda (2010), the theorem follows.

1. To every subset $S \subseteq [n]$, assign a non-negative weight $Y_S \in [1, V)$ for some V that depends on the granularity of our precision, and show that a normalized $d_{\mathrm{TV}}(P, \mathbb{U})$ is equal to $\sum_{S \subseteq [n]} Y_S$.
2. Let k_i be the number of sets $S \subseteq [n]$ for which Y_S lies in $\left[(1+\varepsilon)^{i-1}, (1+\varepsilon)^i\right)$. Then
$$\sum_{i \in [\mathrm{poly}(n)]} k_i \,(1+\varepsilon)^i \approx_\varepsilon \sum_{S \subseteq [n]} Y_S = M \cdot d_{\mathrm{TV}}(P, \mathbb{U})\,,$$
where M is a normalization constant.
3. Reduce the computation of each k_i to $O(1)$ #Knapsack instances.

Open Problems

1. Does there exist an FPTAS for approximating the TV distance between two product distributions?
2. For what other classes of probabilistic models do there exist TV distance approximation schemes?
3. What about other notions of distance or similarity between probabilistic models?

Our Work on arXiv

LaTeX TikZposter