A simple polynomial-time approximation algorithm for the total variation distance between two product distributions

Weiming Feng
University of Edinburgh
Joint work with



Heng Guo
Edinburgh

Mark Jerrum
QMUL

Jiaheng Wang
Edinburgh

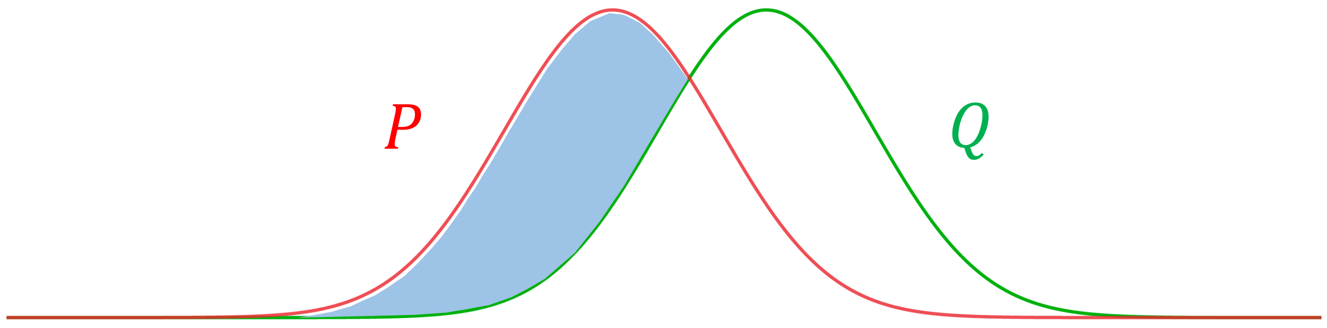
Swiss Winter School on TCS, Zinal, Switzerland

Feb 2nd, 2023

Total variation (TV) distance

Data: two distributions P and Q over state space Ω

Question: how to measure the difference between P and Q



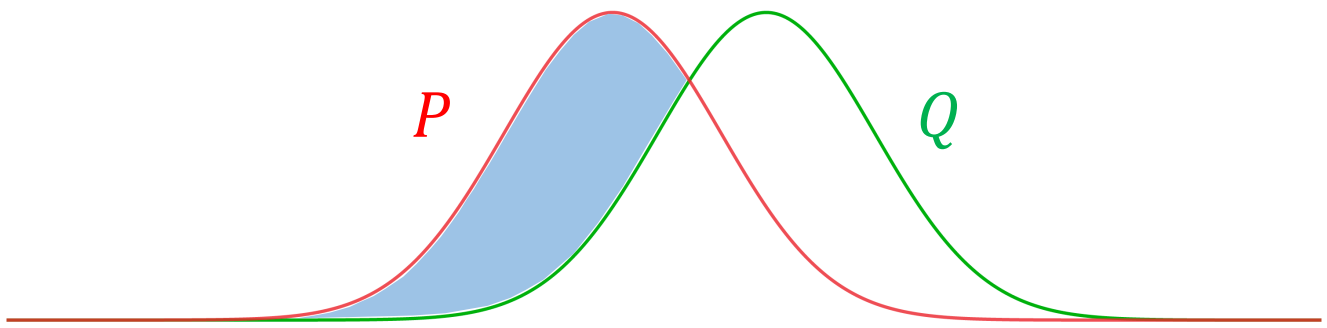
Total variation (TV) distance between P and Q over state space Ω

$$d_{TV}(P,Q) = \frac{1}{2}\sum_{x\in\Omega} |P(x) - Q(x)|$$

Total variation distance

Total variation (TV) distance between P and Q over state space Ω

$$d_{TV}(P,Q) = \frac{1}{2}\sum_{x\in\Omega} |P(x) - Q(x)| = \max_{S\subseteq\Omega} |P(S) - Q(S)|$$



Properties of TV distance

- metric (triangle inequality)
- bounded
- data processing inequality
- various characterisations

Applications of TV distance

- property testing
- Markov chain mixing time
- approximate algorithms
- learning algorithms

Compute TV distance

[Bhattacharyya, Gayen, Meel, Myrisiotis,Pavan, Vinodchandran, 2022]

- **Input:** descriptions of two distributions P, Q over Ω
- **Output:** the total variation distance between P and Q

Trivial algorithm: enumerate all $x \in \Omega$ and add $\frac{1}{2}|P(x) - Q(x)|$ together

Challenge:

- distributions P and Q have ***succinct descriptions***
- $|\Omega|$ can be ***exponentially large*** w.r.t. the size of input

TV distance between two product distributions

$P_1, P_2, \ldots P_n$ and $Q_1, Q_2, \ldots, Q_n$ over finite domain $[s] = \{0,1,\ldots,s-1\}$

P, Q two **product distributions** over **domain** $[s]^n$

$$P = P_1 \times P_2 \times \cdots \times P_n \text{ and } Q = Q_1 \times Q_2 \times \cdots \times Q_n$$

i.e. $$\forall \sigma \in [s]^n, P(\sigma) = \prod_{i=1}^{n} P_i(\sigma_i) \text{ and } Q(\sigma) = \prod_{i=1}^{n} Q_i(\sigma_i)$$

Compute TV distance between two product distributions

[Bhattacharyya, Gayen, Meel, Myrisiotis,Pavan, Vinodchandran, 2022]

- **Input:** distributions $\{P_i, Q_i | 1 \leq i \leq n\}$ specifying P and Q
- **Output:** the total variation distance between P and Q

Input size: $2ns = O(n)$ numbers, each of poly(n) bits **Sample space size of P, Q**: s^n

Theorem [Bhattacharyya, Gayen, Meel, Myrisiotis,Pavan, Vinodchandran, 2022]

Computing TV distance between two **Boolean** $(s = 2)$ product distributions is #P complete.

Approximate TV distance between two product distributions

- **Input:** distributions $\{P_i, Q_i | 1 \leq i \leq n\}$ specifying P and Q
  an error bound $0 < \epsilon < 1$
- **Output:** a random number $\hat{d}$ such that
  $$\Pr\left[(1-\epsilon)d_{TV}(P,Q) \leq \hat{d} \leq (1+\epsilon)d_{TV}(P,Q)\right] \geq 2/3$$

One challenge for approximation

$d_{TV}(P,Q)$ can be **exponentially small**

v.s.

multiplicative approximation error

Theorem [Bhattacharyya, Gayen, Meel, Myrisiotis,Pavan, Vinodchandran, 2022]

There is an **FPRAS** for the TV distance between two **Boolean** product distributions if $\underbrace{\frac{1}{2} \leq P_i(1) \leq 1 \text{ and } 0 \leq Q_{\mathrm{i}}(1) \leq P_i(1) \text{ for all } 1 \leq i \leq n}_{\textbf{additional condition: } \text{a marginal lower bound}}$

FPRAS (Full Poly-time Randomised Approximation Scheme)

An algorithm solves the approximation problem in time poly$(n, 1/\epsilon)$

Open Problem: FPRAS for **general** product distributions

Our results

Main Theorem [F, Guo, Jerrum, Wang, SOSA 2023]

There is an **FPRAS** for the TV distance between two product distributions

- running time $O(n^2/\epsilon^2)$
- work for ***arbitrary finite*** domain
- no extra condition on distributions

TV distance and coupling

- **Distributions**: P and Q over the domain Ω
- **Coupling**: a joint distribution $(X, Y) \in \Omega \times \Omega$ such that $X \sim P$ and $Y \sim Q$

Coupling Lemma (Coupling inequality)

For ***any*** coupling (X, Y) of P and Q,

$$d_{TV}(P, Q) \leq \Pr[X \neq Y]$$

There exists an ***optimal coupling*** of P and Q such that

$$d_{TV}(P, Q) = \Pr[X \neq Y]$$

Greedy coupling between two product distributions

P, Q two **product distributions** over **Boolean domain** $\Omega = [s]^n$

$$P = P_1 \times P_2 \times \cdots \times P_n \text{ and } Q = Q_1 \times Q_2 \times \cdots \times Q_n$$

- **Greedy coupling** $(X, Y) = ((X_1, X_2, \ldots, X_n), (Y_1, Y_2, \ldots, Y_n))$ of P and Q
- Couple each (X_i, Y_i) *independently* using the *optimal coupling* of P_i and Q_i

Non optimal: ∃ product distributions, s.t. $\Pr_{\text{greedy}}[X \neq Y] > d_{TV}(P, Q)$

n-Approximation: ∀ product distributions,

$$d_{TV}(P, Q) \underset{\uparrow}{\leq} \Pr_{\text{greedy}}[X \neq Y] \underset{\uparrow}{\leq} n \cdot d_{TV}(P, Q)$$

Coupling Lemma

Proved by a Union Bound

Property of greedy coupling: greedy coupling and TV distance

$$R = \frac{d_{TV}(P,Q)}{\Pr_{\text{greedy}}[X \neq Y]} \geq \frac{1}{n}$$

Proposition: In greedy coupling, the probability of $X \neq Y$ **is easy to compute**

$$\Pr_{\text{greedy}}[X \neq Y] = 1 - \Pr[X = Y] = 1 - \prod_{i=1}^{n} \left(1 - d_{TV}(P_i, Q_i)\right)$$

Our idea: try to estimate the ratio

$$R = \frac{d_{TV}(P,Q)}{\Pr_{\text{greedy}}[X \neq Y]}$$

$d_{TV}(P,Q)$ can be ***exponentially small*** but the ratio R is **lower bounded by** $1/n$

Our Estimator [F., Guo, Jerrum, Wang, SOSA 2023]

- π: distribution over $[s]^n$ s.t.

$$\forall \sigma \in [s]^n, \qquad \pi(\sigma) = \Pr_{\text{greedy}}[X = \sigma \mid X \neq Y]$$

distribution of X in the greedy coupling conditional on $X \neq Y$

- f: a function $[\mathrm{s}]^n \to \mathbb{R}_{>0}$ s.t.

$$\forall \sigma \in [s]^n, \qquad f(\sigma) = \frac{\Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]}$$

- **Estimator:** $f(\sigma)$ where $\sigma \sim \pi$

Correct expectation

$$\mathbb{E}_{\sigma\sim\pi}[f(\sigma)] = \frac{d_{TV}(P,Q)}{\Pr_{\text{greedy}}[X \neq Y]} = R \geq \frac{1}{n}$$

Low variance

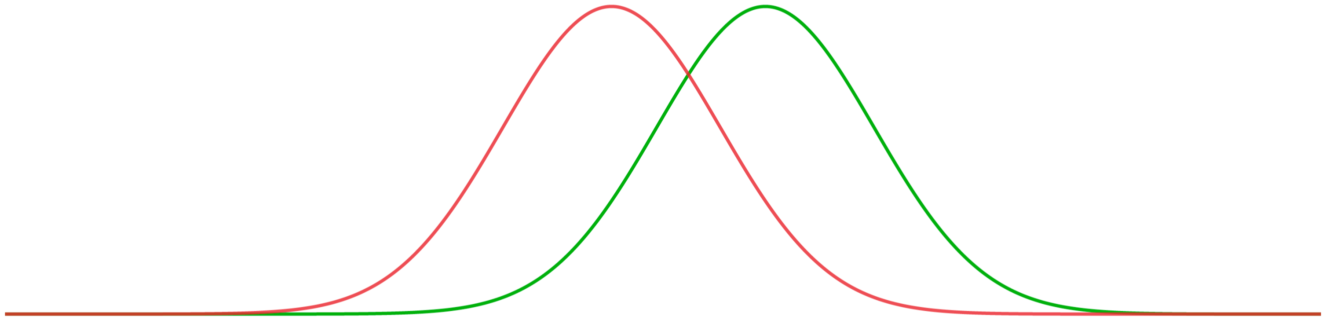
$$\mathrm{Var}_{\sigma\sim\pi}[f(\sigma)] \leq 1$$

Efficient computation

- a random sample of $\sigma \sim \pi$ can be generated in time $O(n)$
- given any $\sigma \in \{0,1\}^n$, $f(\sigma)$ can be computed in time $O(n)$

Summary and open problems

Summary: an FPRAS for the TV distance between two product distributions



Open problems:

- Deterministic approximate algorithm (FPTAS)?
- Beyond the product distributions?

Thanks
Q&A