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

Weiming Feng[1], Heng Guo[1], Mark Jerrum[2], Jiaheng Wang[1]
(1) University of Edinburgh (2) Queen Mary, University of London

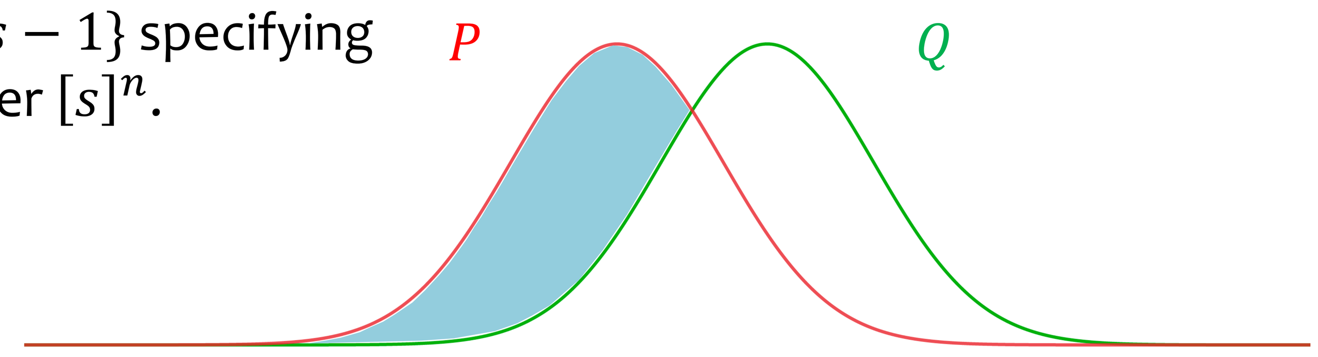
The total variation distance between two product distributions

Input: distributions $P_1, P_2, \dots, P_n$ and $Q_1, Q_2, \dots, Q_n$ over finite domain $[s] = \{0,1,\dots,s-1\}$ specifying two **product distributions** $P = P_1 \times P_2 \times \cdots \times P_n$ and $Q = Q_1 \times Q_2 \times \cdots \times Q_n$ over $[s]^n$.

Output: the **total variation distance (TV distance)** between P, Q

$$d_{TV}(P,Q) = \frac{1}{2}\sum_{x\in[s]^n} |P(x) - Q(x)| = \max_{A\subseteq[s]^n} |P(A) - Q(A)|.$$

Challenge: the size of sample space of P, Q is **exponentially large** s^n.



Background & previous results [Bhattacharyya, Gayen, Meel, Myrisiotis, Pavan, Vinodchandran, 2022]

Hardness: the exact computation of the TV distance between two product distributions is **#P-complete**.

Approximate the TV distance: Given P, Q and an error bound $0 < \epsilon < 1$, an **FPRAS** outputs a random $\hat{d}$ in time poly$(n, 1/\epsilon)$ such that

$$\Pr\left[(1-\epsilon)d_{TV}(P,Q) \le \hat{d} \le (1+\epsilon)d_{TV}(P,Q)\right] \ge \frac{2}{3}.$$

Algorithm: there is an FPRAS if $s = 2$ (Boolean domain) and $\frac{1}{2} \le P_i(1) < 1$ and $0 < Q_i(1) \le P_i(1)$ for all $1 \le i \le n$.

Our results [Feng, Guo, Jerrum, Wang, SOSA 2023]

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

- $O(n^2/\epsilon^2)$-time assuming the cost of each arithmetic operation is $O(1)$;
- each operation acts on poly(n)-bit numbers if input par. has poly(n) bits.

Open problems

- **Deterministic** approximate algorithm (FPTAS)?
- FPTAS/FPRAS **beyond** the product distribution?

TV distance and coupling

A **Coupling** of two distribution P, Q over Ω is a pair of **joint random variables**

$$(X, Y) \in \Omega\times\Omega \text{ such that } X \sim P \text{ and } Y \sim Q$$

Coupling lemma: for any coupling (X, Y) of P, Q,

$$d_{TV}(P,Q) \le \Pr_{\text{coupling}}[X \ne Y],$$

and there exists **an optimal coupling** of P, Q such that

$$d_{TV}(P,Q) = \Pr_{\text{opt}}[X \ne Y].$$

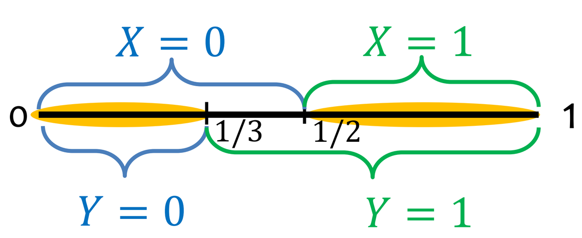
Examples of coupling

Distributions: $P(0) = P(1) = \frac{1}{2}$ and $Q(0) = \frac{1}{3}, Q(1) = \frac{2}{3}$.

Ind. Coupling: sample $X \sim P, Y \sim Q$ independently

$$\Pr_{\text{ind}}[X \ne Y] = 1/2.$$

Opt. Coupling: sample real $r \in [0,1]$ uniformly at random $X = 0$ iff $r \le 1/2$ and $Y = 0$ iff $r \le 1/3$.



$$\Pr_{\text{opt}}[X \ne Y] = 1/6$$

Greedy coupling for product distributions

Product distributions: $P = P_1 \times P_2 \times \cdots \times P_n$ and $Q = Q_1 \times Q_2 \times \cdots \times Q_n$.

Greedy coupling: couple each (P_i, Q_i) **optimally** and **independently**.

Example: *Boolean distributions* $P, Q \in \{0,1\}^n$

- sample $r_i \in [0,1]$ uniformly and ind. for all i;
- $X_i = 0$ iff $r_i \le P_i(0)$ and $Y_i = 0$ iff $r_i \le Q_i(0)$;
- $X = (X_1, X_2, \dots, X_n)$ and $Y = (Y_1, Y_2, \dots, Y_n)$.

Properties of the greedy coupling

Non-optimal: Ex $P = P_1 \times P_2$ and $Q = Q_1 \times Q_2$ s.t. $\forall i$

$$P_i(0) = P_i(1) = \frac{1}{2} \text{ and } Q_i(0) = \frac{1}{2} - \delta, Q_i(1) = \frac{1}{2} + \delta.$$

$$\Pr[X \ne Y] = \delta(2-\delta) > \delta(1+\delta) = d_{TV}(P,Q).$$

Poly(n)-approximation: for any product distributions,

$$d_{TV}(P,Q) \le \Pr_{\text{greedy}}[X \ne Y] \le n \cdot d_{TV}(P,Q).$$

Our idea: estimate the ratio

Fact: Prob. of $X \ne Y$ in greedy coupling is **easy to compute**

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

Lemma: There is FPRAS for the ratio

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

Estimator for the ratio R:

- π: distribution of X in the **greedy coupling conditional on** $X \ne Y$

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

- f: a function $[s]^n \to \mathbb{R}_{>0}$ such that $\forall \sigma \in [s]^n$

$$f(\sigma) = \frac{\Pr_{\text{opt}}[X = \sigma \wedge X \ne Y]}{\Pr_{\text{greedy}}[X = \sigma \wedge X \ne Y]} = \frac{\max\{0, P(\sigma) - Q(\sigma)\}}{\Pr_{\text{greedy}}[X = \sigma \wedge X \ne Y]}.$$

- Estimator: $W = f(\sigma)$ where $\sigma \sim \pi$.

Properties of the estimator

- **Correct** expectation

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

- **Low** variance

$$\text{Var}_{\sigma\sim\pi}[f(\sigma)] \le 1.$$

- **Efficient** computation
  - sample $\sigma \sim \pi$ in time $O(n)$;
  - given any $\sigma \in \{0,1\}^n$, compute $f(\sigma)$ in time $O(n)$.

Algorithm

- draw $\sigma_1, \sigma_2, \dots \sigma_m \sim \pi$ for $m = (n/\epsilon^2)$;
- return the average $\hat{R} = \frac{1}{m}\sum_{i=1}^{m} f(\sigma_i)$.

Correctness: Chebyshev ineq. + $\text{Var}[f] \le 1$ + $\mathbb{E}[f] \ge \frac{1}{n}$.

Efficiency: efficient computation property.