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

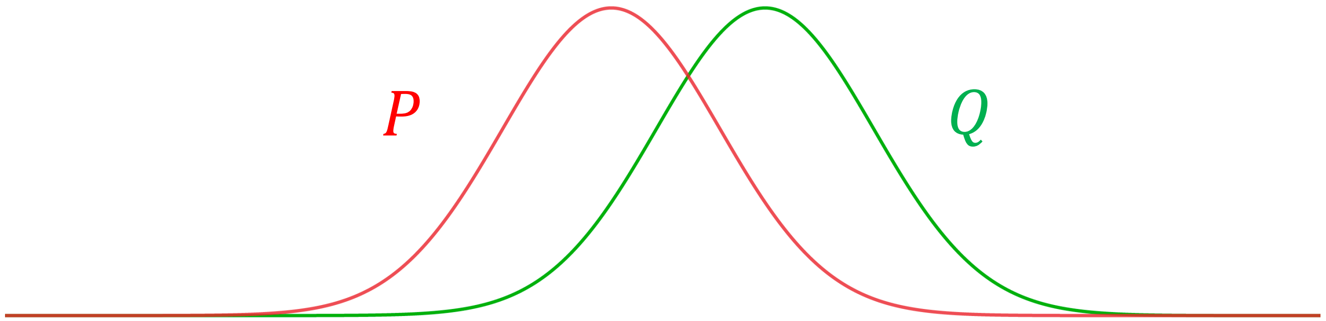
Peking University, Beijing, China

Feb 21st, 2023

Difference between two distributions

Data: two distributions P and Q over state space Ω

Question: how to measure the difference between P and Q

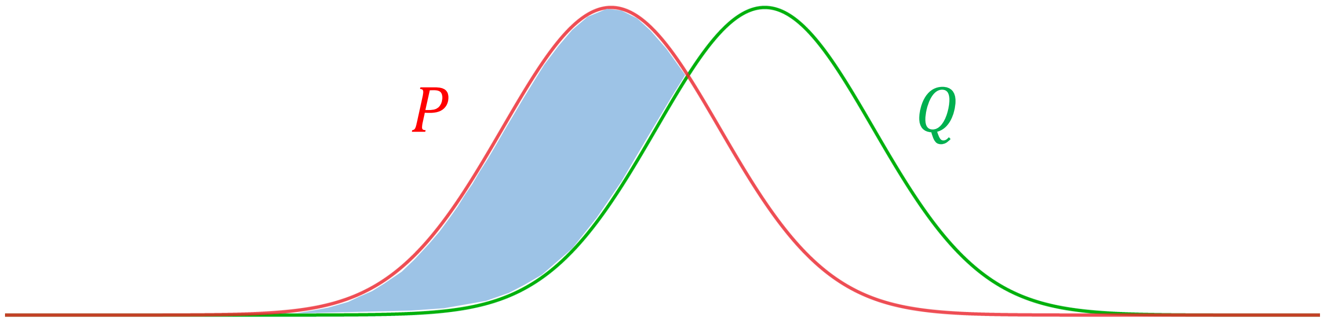


- Total variation distance (TV distance): $d_{TV}(P,Q) = \frac{1}{2}\sum_{x\in\Omega}|P(x) - Q(x)|$
- KL-divergence (relative entropy): $D_{KL}(P||Q) = \sum_{x\in\Omega} P(x)\log\frac{P(x)}{Q(x)}$
- χ^2-divergence: $D_{\chi^2}(P||Q) = \left(\sum_{x\in\Omega}\frac{P^2(X)}{Q(x)}\right) - 1$

Total variation (TV) 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

Examples: probabilistic graphical models, spin systems.

TV distance between two product distributions

Distributions $P_1, P_2, \ldots P_n$ and $Q_1, Q_2, \ldots, Q_n$ over $\{0,1\}$

- P_i: distribution over $\{0,1\}$ such that $P_i(1) = p_i$ and $P_i(0) = 1 - p_i$
- Q_i: distribution over $\{0,1\}$ such that $Q_i(1) = q_i$ and $Q_i(0) = 1 - q_i$

Product distributions P and Q over $\{0,1\}^n$

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

Random sample $X = (X_1, X_2, \ldots, X_n) \sim P$

$X \in \{0,1\}^n$: n-dimensional random vector

$X_i \in \{0,1\}$: independent sample from P_i

$$\forall\, X \in \{0,1\}^n, \qquad P(X) = \prod_{i=1}^{n} P_i(X_i)$$

TV distance between two product distributions

Distributions $P_1, P_2, \dots P_n$ and $Q_1, Q_2, \dots, Q_n$ over $\{0,1\}$

- P_i: distribution over $\{0,1\}$ such that $P_i(1) = p_i$ and $P_i(0) = 1 - p_i$
- Q_i: distribution over $\{0,1\}$ such that $Q_i(1) = q_i$ and $Q_i(0) = 1 - q_i$

Product distributions P and Q over $\{0,1\}^n$

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

Compute TV distance between two Boolean product distributions

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

- **Input:** vectors $(p_1, p_2, \dots, p_n)$ and $(q_1, q_2, \dots, q_n)$ specifying P and Q
- **Output:** the total variation distance between P and Q

Input size: $2n$ numbers, each has $\mathrm{poly}(n)$ bits

Sample space size: 2^n

TV distance between two product distributions

Finite domain $[\mathrm{s}] = \{0,1,\ldots,s-1\}$ with constant s

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

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

- P_i, Q_i distributions over $[s]$

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

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 number $\hat{d}$ such that $(1-\epsilon)d_{TV}(P,Q) \leq \hat{d} \leq (1+\epsilon)d_{TV}(P,Q)$

FPRAS (Full Poly-time Randomised Approximation Scheme)

A randomised algorithm outputs a ***random*** $\hat{d}$ in time poly$(n, 1/\epsilon)$

$$\Pr\left[(1-\epsilon)d_{TV}(P,Q) \leq \hat{d} \leq (1+\epsilon)d_{TV}(P,Q)\right] \geq 2/3$$

Previous results

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

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

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

Our results

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

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

- work for ***arbitrary finite*** domain
- no extra condition on distributions
- simple algorithm

In this talk, for simplicity,
I always use ***Boolean*** (s=2) distribution as
an ***example*** to explain our technique

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

Let $s \geq 2$ be a constant. There is an algorithm such that

- *Input*: two product distributions P, Q over $[s]^n$
- *Output*: a random $\hat{d}$ *that* ϵ-approximates $d_{TV}(P,Q)$ with prob. $\geq \frac{2}{3}$
- *Running time*: $O\left(\frac{n^2}{\epsilon^2}\right)$ assuming the cost of each arithmetic operation is $O(1)$
- *Bit length*: each arithmetic operator works on two numbers with $\text{poly}(n)$ bits
  if each input parameter has $\text{poly}(n)$ bits.

A natural estimators for TV distance [BGMV20]

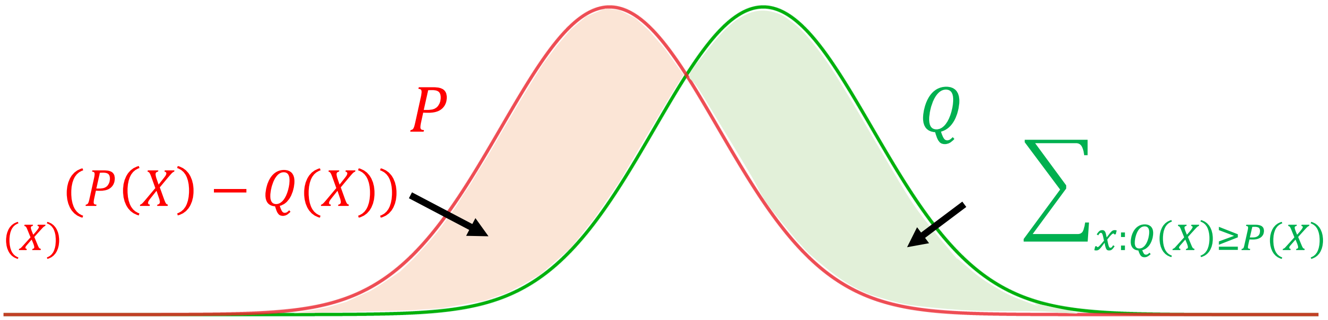
- Draw a random sample $X \sim P$
- Compute the estimator

$$W = W(X) = \frac{\max\{0, P(X) - Q(X)\}}{P(X)}$$

Unbiased Estimator

$$\mathbb{E}[W] = \sum_x P(x)W(x) = \sum_{x:P(X)\geq Q(X)} (P(X) - Q(X))$$

$$= \frac{1}{2}\sum_x |P(X) - Q(X)| = d_{TV}(P, Q).$$



$$\sum_{x:P(X)\geq Q(X)} (P(X) - Q(X))$$

$$\sum_{x:Q(X)\geq P(X)} (Q(X) - P(X))$$

A natural estimators for TV distance [BGMV20]

- Draw a random sample $X \sim P$
- Compute the estimator

$$W = W(X) = \frac{\max\{0, P(X) - Q(X)\}}{P(X)}$$

Unbiased Estimator: $\mathbb{E}[W] = d_{TV}(P, Q)$ **Boundedness:** $\forall x,\ 0 \le W(x) \le 1$

- Sample $W_1, W_2, W_3, \dots W_m$ independently for $m = \text{poly}(n, 1/\epsilon)$;
- Output the average $\widehat{W} = (W_1 + W_2 + \cdots + W_m)/m$

Good for additive error (Hoeffding's bound): $d_{TV}(P, Q) - \epsilon \le \widehat{W} \le d_{TV}(P, Q) + \epsilon$

NO, because $d_{TV}(P, Q)$ can be $\exp(-\text{poly}(n))$

Relative error ? $(1 - \epsilon) d_{TV}(P, Q) \le \widehat{W} \le (1 + \epsilon) d_{TV}(P, Q)$

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 $Y \sim Q$

x	0	1
$P(x)$	$1/2$	$1/2$

y	0	1
$Q(y)$	$1/3$	$2/3$

Example: Independent Coupling

Sample $X \sim P$ and $Y \sim Q$ independently

$$\Pr[X \neq Y] = \frac{1}{2}$$

Can we make this prob. smaller?

TV distance and coupling

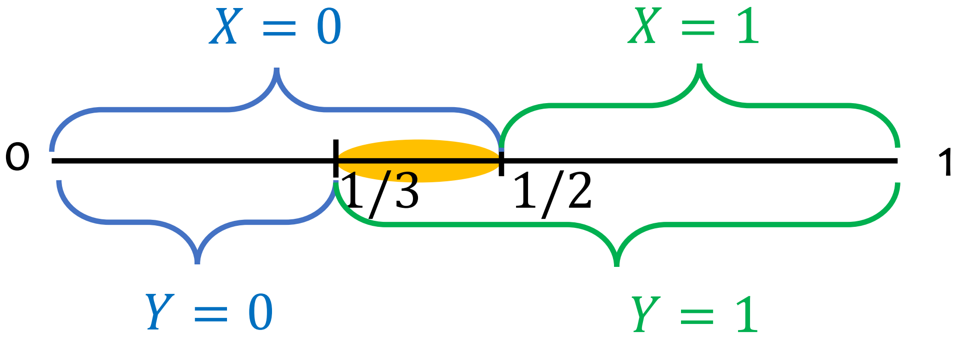
- **Distributions**: P and Q over the domain Ω
- **Coupling**: a joint distribution $(X, Y) \in \Omega \times \Omega$ such that $X \sim P$ and $Y \sim Q$

x	0	1
$P(x)$	$1/2$	$1/2$

y	0	1
$Q(y)$	$1/3$	$2/3$

Example: Optimal Coupling

- Sample $r \in (0,1)$ uniformly at random
- Let $X = 0$ iff $r < P(0) = 1/2$
- Let $Y = 0$ iff $r < Q(0) = 1/3$



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 $Y \sim Q$

x	0	1
$P(x)$	$1/2$	$1/2$

y	0	1
$Q(y)$	$1/3$	$2/3$

Example: Optimal Coupling

- Sample $r \in (0,1)$ uniformly at random
- Let $X = 0$ iff $r < P(0) = 1/2$
- Let $Y = 0$ iff $r < Q(0) = 1/3$

$$\Pr[X \neq Y] = \frac{1}{6} = d_{TV}(P, Q)$$

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** $\Omega = \{0,1\}^n$

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

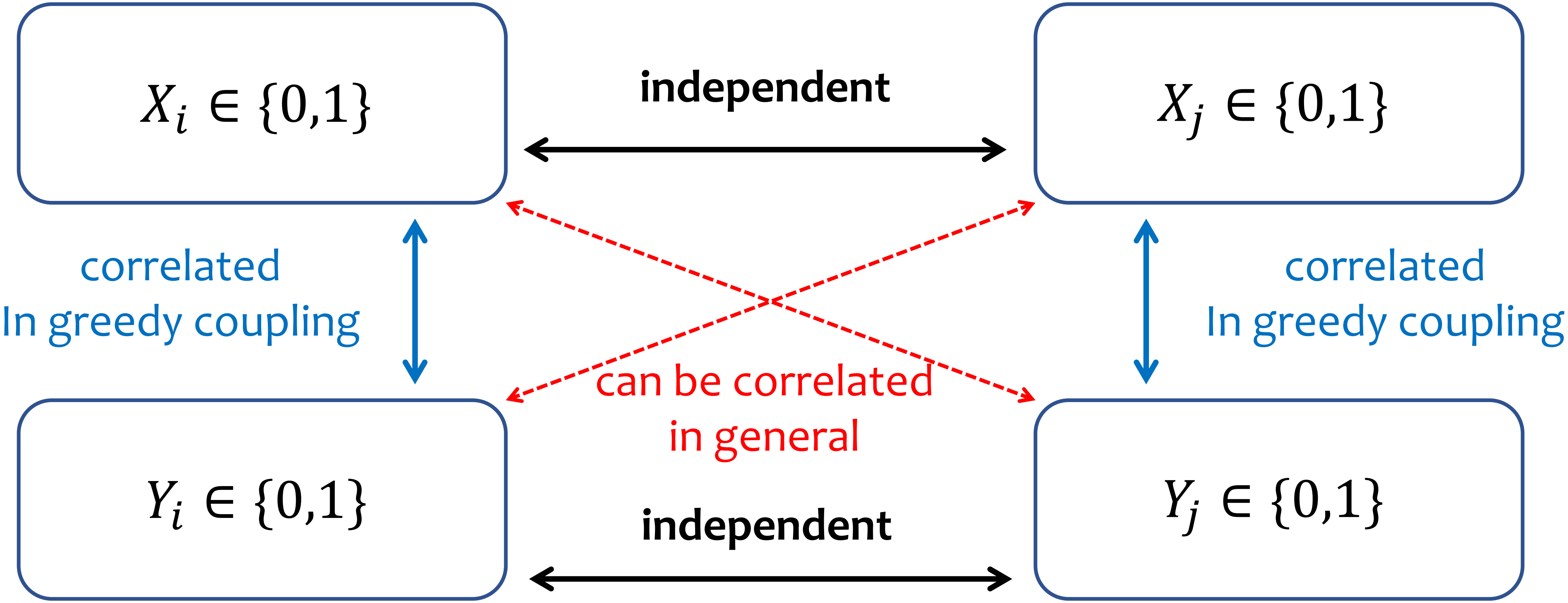
- **Greedy coupling** $(X, Y) = ((X_1, X_2, \ldots, X_n), (Y_1, Y_2, \ldots, Y_n))$ of P and Q
- Sample each (X_i, Y_i) *independently* for the *optimal coupling* of P_i and Q_i

- Sample real numbers $r_i \in [0,1]$ *uniformly* and *independently* for all $1 \leq i \leq n$
- For each $1 \leq i \leq n$, couple (X_i, Y_i) *optimally* by

$$X_i = \begin{cases} 0 & \text{if } r_i < P_i(0) \\ 1 & \text{if } r_i \geq P_i(0) \end{cases} \qquad Y_i = \begin{cases} 0 & \text{if } r_i < Q_i(0) \\ 1 & \text{if } r_i \geq Q_i(0) \end{cases}$$

- Output random vectors $X = (X_1, X_2, \ldots, X_n)$ and $Y = (Y_1, Y_2, \ldots, Y_n)$

Proposition: For product distributions, the **greedy coupling** is **not optimal**



$X_i \in \{0,1\}$

independent

$X_j \in \{0,1\}$

correlated
In greedy coupling

can be correlated
in general

correlated
In greedy coupling

$Y_i \in \{0,1\}$

independent

$Y_j \in \{0,1\}$

It may be possible to utilise the ***correlation in the middle*** to get a better coupling

Proposition: For product distributions, the **greedy coupling** is **not optimal**

- $P = P_1 \times P_2 \times \cdots \times P_n$, where for each $1 \le i \le n$, $P_i(1) = P_i(0) = \frac{1}{2}$
- $Q = Q_1 \times Q_2 \times \cdots \times Q_n$, where for each $1 \le i \le n$, $Q_i(1) = \frac{1}{2} + \delta$ and $Q_i(0) = \frac{1}{2} - \delta$
- Suppose $\delta = \exp(-\Omega(n))$ is small

Total variation distance between P and Q

(Pinsker's ineq)

$$d_{TV}(P, Q) \le \sqrt{D_{KL}(P||Q)} = \sqrt{\sum_{i=1}^{n} D_{KL}(P_i||Q_i)} = O(\delta\sqrt{n})$$

Greedy coupling (X, Y) of P and Q

$$\Pr[X \neq Y] = 1 - (1 - \delta)^n = \Omega(\delta n)$$

Greedy coupling can be $\Omega(\sqrt{n})$-times **worse** than the **optimal coupling**

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

Proposition: For product distributions, the **greedy coupling** is **not optimal,** but the **greedy coupling** **cannot be too bad**

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

Proposition: For product distributions, the **greedy coupling** is **not optimal,** but the **greedy coupling** **cannot be too bad**

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

$$\Pr_{\text{greedy}}[X \neq Y] = \Pr[\exists i, X_i \neq Y_i] \underset{\substack{\uparrow \\ \textbf{union bound}}}{\leq} \sum_{i=1}^{n} \Pr[X_i \neq Y_i] \underset{\substack{\uparrow \\ \textbf{coupling lemma}}}{=} \sum_{i=1}^{n} d_{TV}(P_i, Q_i)$$

Proposition: For product distributions, the **greedy coupling** is **not optimal,** but the **greedy coupling** **cannot be too bad**

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

$$\Pr_{\text{greedy}}[X \neq Y] = \Pr[\exists i, X_i \neq Y_i] \underset{\substack{\uparrow \\ \text{union bound}}}{\le} \sum_{i=1}^{n} \Pr[X_i \neq Y_i] \underset{\substack{\uparrow \\ \text{coupling lemma}}}{=} \sum_{i=1}^{n} d_{TV}(P_i, Q_i)$$

$$d_{TV}(P,Q) \underset{\substack{\uparrow \\ \text{coupling lemma}}}{=} \Pr_{\text{opt}}[X \neq Y] \ge \Pr_{\text{opt}}[X_i \neq Y_i]$$

Proposition: For product distributions, the **greedy coupling** is **not optimal,** but the **greedy coupling** **cannot be too bad**

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

$$\Pr_{\text{greedy}}[X \ne Y] = \Pr[\exists i, X_i \ne Y_i] \underset{\substack{\uparrow \\ \textbf{union bound}}}{\le} \sum_{i=1}^{n} \Pr[X_i \ne Y_i] \underset{\substack{\uparrow \\ \textbf{coupling lemma}}}{=} \sum_{i=1}^{n} d_{TV}(P_i, Q_i)$$

$$d_{TV}(P,Q) \underset{\substack{\uparrow \\ \textbf{coupling lemma}}}{=} \Pr_{\text{opt}}[X \ne Y] \ge \Pr_{\text{opt}}[X_i \ne Y_i] \underset{\substack{\uparrow \\ \textbf{coupling lemma}}}{\ge} d_{TV}(P_i, Q_i)$$

Proposition: For product distributions, the **greedy coupling** is **not optimal,** but the **greedy coupling cannot be too bad**

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

$$\Pr_{\text{greedy}}[X \neq Y] = \Pr[\exists i, X_i \neq Y_i] \le \sum_{i=1}^{n} \Pr[X_i \neq Y_i] = \sum_{i=1}^{n} d_{TV}(P_i,Q_i) \le n \cdot d_{TV}(P,Q)$$

union bound **coupling lemma**

$$d_{TV}(P,Q) = \Pr_{\text{opt}}[X \neq Y] \ge \Pr_{\text{opt}}[X_i \neq Y_i] \ge d_{TV}(P_i,Q_i)$$

coupling lemma **coupling lemma**

Proposition: **greedy coupling** and **TV distance**

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

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**

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

Compared with $d_{TV}(P,Q)$, the ratio R is lower bounded by $1/n$

Proposition: **greedy coupling** and **TV distance**

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

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

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

Lemma [F., Guo, Jerrum, Wang, SOSA 2023]

There is an algorithm that outputs $\hat{R}$ *in time* $O(n^2/\epsilon^2)$ *such that*

$$\Pr\left[(1-\epsilon)R \le \hat{R} \le (1+\epsilon)R\right] \ge \frac{2}{3}, \qquad \text{where } R = \frac{d_{TV}(P,Q)}{\Pr_{\text{greedy}}[X \neq Y]}$$

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

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

$$\forall \sigma \in \{0,1\}^n, \qquad \pi(\sigma) = \Pr_{\text{greedy}}[X = \sigma \mid X \neq Y]$$

- f: a function $\{0,1\}^n \to \mathbb{R}_{>0}$ such that

$$\forall \sigma \in \{0,1\}^n, \qquad f(\sigma) = \frac{\Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]}$$

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

Lemma: for ***any optimal coupling*** (X, Y) of P, Q,

$$\forall \sigma \in \{0,1\}^n, \ \Pr_{\text{opt}}[X = Y = \sigma] = \min\{P(\sigma), Q(\sigma)\}$$

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

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

$$\forall \sigma \in \{0,1\}^n, \qquad \pi(\sigma) = \Pr_{\text{greedy}}[X = \sigma \mid X \neq Y]$$

- f: a function $\{0,1\}^V \to \mathbb{R}_{>0}$ such that

$$\forall \sigma \in \{0,1\}^n, \qquad f(\sigma) = \frac{\Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]} = \frac{P(\sigma) - \min\{P(\sigma), Q(\sigma)\}}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]}$$

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

Lemma: for ***any*** optimal coupling (X, Y) of P, Q,

$$\forall \sigma \in \{0,1\}^n, \ \Pr_{\text{opt}}[X = Y = \sigma] = \min\{P(\sigma), Q(\sigma)\}$$

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

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

$$\forall \sigma \in \{0,1\}^n, \qquad \pi(\sigma) = \Pr_{\text{greedy}}[X = \sigma \mid X \neq Y]$$

- f: a function $\{0,1\}^V \to \mathbb{R}_{>0}$ such that

$$\forall \sigma \in \{0,1\}^n, \qquad f(\sigma) = \frac{\Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]} = \frac{\max\{0, P(\sigma) - Q(\sigma)\}}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]}$$

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

Lemma: for ***any*** optimal coupling (X, Y) of P, Q,

$$\forall \sigma \in \{0,1\}^n, \ \Pr_{\text{opt}}[X = Y = \sigma] = \min\{P(\sigma), Q(\sigma)\}$$

Remark about the lemma

- For any coupling (X, Y), for all $\sigma \in \{0,1\}^n$, $\Pr[X = Y = \sigma] \leq \min\{P(\sigma), Q(\sigma)\}$
- The optimal coupling maximise the $\Pr_{\text{opt}}[X = Y]$ by

maximise $\Pr_{\text{opt}}[X = Y = \sigma]$ ***for all*** $\sigma \in \{0,1\}^n$ ***at the same time***

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

$$f(\sigma) = \frac{\Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]} = \frac{\max\{0, P(\sigma) - Q(\sigma)\}}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]}, \qquad \text{where } \sigma \sim \pi$$

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

- **Correct expectation with lower bound**

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

- **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 $O(n)$

draw $O(\frac{n}{\epsilon^2})$ ind. samples $\sigma \sim \pi$
compute the average of $f(\sigma)$

Chebyshev's inequality

FPRAS for estimating R

Proof: Correct expectation

$$\mathbb{E}_{\sigma\sim\pi}[f(\sigma)] = \sum_{\sigma} \pi(\sigma) f(\sigma) \quad \text{(sum over all } \sigma \in \{0,1\}^n \text{ with } \pi(\sigma) > 0\text{)}$$

Proof: Correct expectation

$$\mathbb{E}_{\sigma\sim\pi}[f(\sigma)] = \sum_{\sigma} \pi(\sigma) f(\sigma)$$

(sum over all $\sigma \in \{0,1\}^n$ with $\pi(\sigma) > 0$)

$$= \sum_{\sigma} \Pr_{\text{greedy}}[X = \sigma \mid X \neq Y] \frac{\Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]}$$

by definitions of π and f

Proof: Correct expectation

$$\mathbb{E}_{\sigma\sim\pi}[f(\sigma)] = \sum_{\sigma} \pi(\sigma) f(\sigma)$$

(sum over all $\sigma \in \{0,1\}^n$ with $\pi(\sigma) > 0$)

$$= \sum_{\sigma} \Pr_{\text{greedy}}[X = \sigma \mid X \neq Y] \frac{\Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]}$$

$$= \sum_{\sigma} \frac{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]}{\Pr_{\text{greedy}}[X \neq Y]} \cdot \frac{\Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]}$$

by the definition of conditional distribution

Proof: Correct expectation

$$\mathbb{E}_{\sigma\sim\pi}[f(\sigma)] = \sum_{\sigma} \pi(\sigma) f(\sigma)$$ (sum over all $\sigma \in \{0,1\}^n$ with $\pi(\sigma) > 0$)

$$= \sum_{\sigma} \underset{\text{greedy}}{\Pr}[X = \sigma \mid X \neq Y] \, \frac{\underset{\text{opt}}{\Pr}[X = \sigma \wedge X \neq Y]}{\underset{\text{greedy}}{\Pr}[X = \sigma \wedge X \neq Y]}$$

$$= \sum_{\sigma} \frac{\cancel{\underset{\text{greedy}}{\Pr}[X = \sigma \wedge X \neq Y]}}{\underset{\text{greedy}}{\Pr}[X \neq Y]} \cdot \frac{\underset{\text{opt}}{\Pr}[X = \sigma \wedge X \neq Y]}{\cancel{\underset{\text{greedy}}{\Pr}[X = \sigma \wedge X \neq Y]}}$$

$$= \frac{1}{\underset{\text{greedy}}{\Pr}[X \neq Y]} \sum_{\sigma} \underset{\text{opt}}{\Pr}[X = \sigma \wedge X \neq Y]$$

Proof: Correct expectation

$$\sum_{\sigma:\pi(\sigma)>0} \Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]$$

(sum over all $\sigma \in \{0,1\}^n$ with $\pi(\sigma) > 0$)

If the summation takes over all $\sigma \in \{0, 1\}^n$, then

$$\sum_{\sigma\in\{0,1\}^n} \Pr_{\text{opt}}[X = \sigma \wedge X \neq Y] = \Pr_{\text{opt}}[X \neq Y]$$

➡

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

QED

Proof: Correct expectation

Lemma: The optimal coupling maximise the $\Pr_{\text{opt}}[X = Y]$ by

maximise $\Pr_{\text{opt}}[X = Y = \sigma]$ ***for all*** $\sigma \in \{0,1\}^n$ ***at the same time***

Corollary: for any $\sigma \in \{0,1\}^n$, it holds that

$$\Pr_{\text{greedy}}[X = Y = \sigma] \leq \Pr_{\text{opt}}[X = Y = \sigma]$$

$$\Longrightarrow \quad \Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y] \geq \Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]$$

$$\boxed{\pi(\sigma) = 0} \Longleftrightarrow \boxed{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y] = 0} \xrightarrow{\text{By corollary}} \boxed{\Pr_{\text{opt}}[X = \sigma \wedge X \neq Y] = 0}$$

$$\sum_{\sigma:\pi(\sigma)>0} \Pr_{\text{opt}}[X = \sigma \wedge \ X \neq Y] = \sum_{\sigma \in \{0,1\}^n} \Pr_{\text{opt}}[X = \sigma \wedge \ X \neq Y] = \Pr_{\text{opt}}[X \neq Y]$$

Proof: Low variance

$$f(\sigma) = \frac{\Pr_{\text{opt}}[X = \sigma \wedge \ X \neq Y]}{\Pr_{\text{greedy}}[X = \sigma \wedge \ X \neq Y]}$$

Property: for any σ with $\pi(\sigma) > 0$,

$$0 \leq f(\sigma) \leq 1$$

Corollary: for any $\sigma \in \{0,1\}^n$, it holds that

$$\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y] \geq \Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]$$

$$\boxed{\forall \sigma, 0 \leq f(\sigma) \leq 1} \Rightarrow \boxed{|f(\sigma) - \mathbb{E}_\pi[f]| \leq 1} \Rightarrow \boxed{\mathrm{Var}_\pi[f] = \mathbb{E}_\pi[|f - \mathbb{E}_\pi[f]|^2] \leq 1}$$

Proof: Efficient computation

Task I: sampling problem

Draw a random sample $\sigma \in \{0,1\}^n$ from the distribution π

- π: the distribution of X in greedy coupling (X, Y) conditional on $X \neq Y$

Task II: computational problem

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

Proof: Efficient computation

Task I: sampling problem

Draw a random sample $\sigma \in \{0,1\}^n$ from the distribution π

- π: the distribution of X in greedy coupling (X, Y) conditional on $X \neq Y$

Task II: computational problem

Given σ, compute $f(\sigma) = \dfrac{\Pr_{\text{opt}}[X=\sigma \wedge X \neq Y]}{\Pr_{\text{greedy}}[X=\sigma \wedge X \neq Y]} = \dfrac{\max\{0, P(\sigma) - Q(\sigma)\}}{\Pr_{\text{greedy}}[X=\sigma \wedge X \neq Y]}$

Task I: sampling problem

Draw a random sample $\sigma \in \{0,1\}^n$ from the distribution π

- π: the distribution of X in greedy coupling (X, Y) conditional on $X \neq Y$

Challenge: π over $\{0,1\}^n$ is ***not*** a product distribution

- The greedy coupling is a **product distribution**

  all pairs (X_i, Y_i) are mutually independent

- The condition $X \neq Y$ is **not complicated**

$$X \neq Y \iff \neg(X = Y) \iff \neg(\wedge_{i=1}^{n} (X_i = Y_i))$$

Proof: Efficient computation

Draw random sample $\sigma \in \{0,1\}^n$ for the distribution π

```
- **For** each $i$ for 1 to $n$ **do**
      sample $\sigma_i$ for $\pi$ conditional on $\sigma_1, \sigma_2, \dots, \sigma_{i-1}$
- **Return** $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$
```

Fix $\sigma_1, \sigma_2, \dots, \sigma_{i-1} \in \{0,1\}$, sample $\sigma_i \in \{0,1\}$ according to

$$\Pr[\sigma_i = 0] = \Pr_{X \sim \pi}[\, X_i = 0 \mid X_1 = \sigma_1 \wedge X_2 = \sigma_2 \wedge \cdots \wedge X_{i-1} = \sigma_{i-1} \,]$$

$$\Pr[\sigma_i = 1] = \Pr_{X \sim \pi}[\, X_i = 1 \mid X_1 = \sigma_1 \wedge X_2 = \sigma_2 \wedge \cdots \wedge X_{i-1} = \sigma_{i-1} \,]$$

Proof: Efficient computation

Draw random sample $\sigma \in \{0,1\}^n$ for the distribution π

- **For** each i for 1 to n **do**
    - sample σ_i for π conditional on $\sigma_1, \sigma_2, \dots, \sigma_{i-1}$
- **Return** $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$

Fix $\sigma_1, \sigma_2, \dots, \sigma_{i-1} \in \{0,1\}$, sample $\sigma_i \in \{0,1\}$ according to

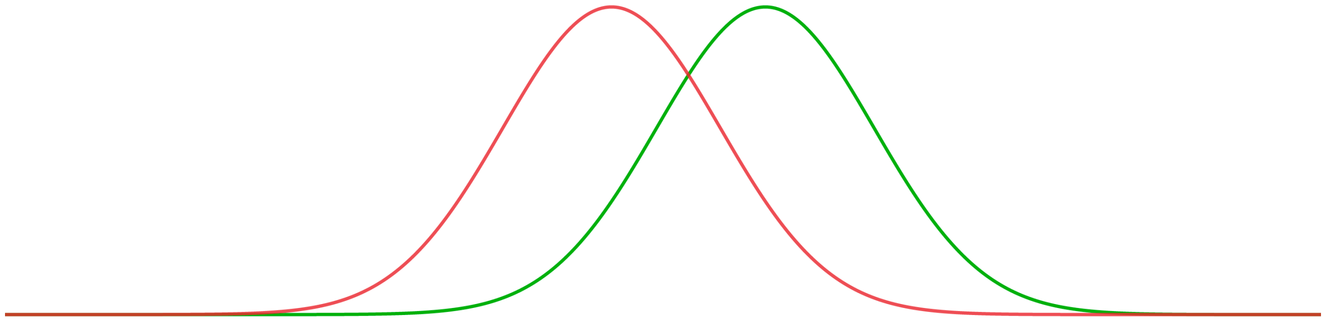
$$\Pr[\sigma_i = 0] = \underset{\text{greedy}}{\Pr}[\,X_i = 0 \mid X_1 = \sigma_1 \wedge X_2 = \sigma_2 \wedge \cdots \wedge X_{i-1} = \sigma_{i-1} \wedge X \neq Y\,]$$

$$\Pr[\sigma_i = 1] = \underset{\text{greedy}}{\Pr}[\,X_i = 1 \mid X_1 = \sigma_1 \wedge X_2 = \sigma_2 \wedge \cdots \wedge X_{i-1} = \sigma_{i-1} \wedge X \neq Y\,]$$

The conditional marginal distribution can be computed ***efficiently***

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