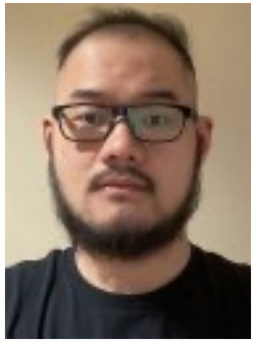
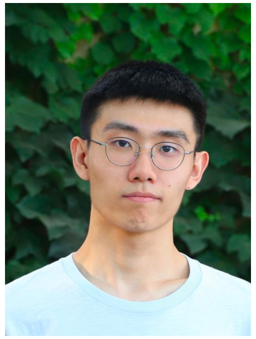
Approximating the total variation distance between two product distributions

Weiming Feng (UC Berkeley)

Based on joint works with







Heng Guo
Edinburgh

Mark Jerrum
QMUL

Jiaheng Wang
Edinburgh

Tianren Liu
Peking

Liqiang Liu
Peking

Randomised Algorithm (Heng Guo, Mark Jerrum, Jiaheng Wang)

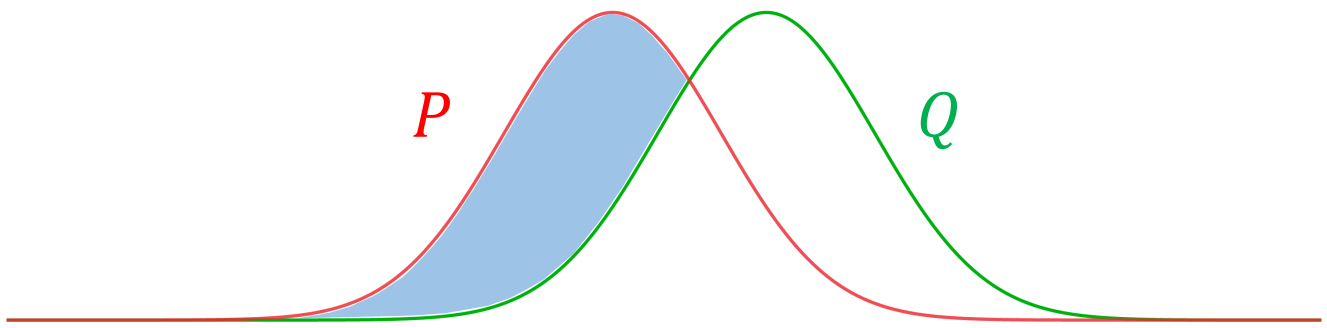
Deterministic Algorithm (Tianren Liu, Liqiang Liu)

Shonan Meeting: MCMC 2.0 Shonan, Japan, Sep 7, 2023

Total Variation distance

Total variation (TV) distance between $\mathbb{P}$ and $\mathbb{Q}$ over state space Ω

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



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 $\mathbb{P}, \mathbb{Q}$ over Ω
- **Output:** the total variation distance between $\mathbb{P}$ and $\mathbb{Q}$

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

Challenge:$\mathbb{P}$ and $\mathbb{Q}$ have ***succinct descriptions***

- $|\Omega|$ can be ***exponentially large*** w.r.t. the size of input

Examples: probabilistic graphical models, spin systems.

Product distribution

Product distribution $\mathbb{P}$ over $[s]^n$

$$\mathbb{P} = \mathbb{P}_1 \times \mathbb{P}_2 \times \cdots \times \mathbb{P}_n$$

$$\forall X, \mathbb{P}(X) = \prod_{i=1}^{n} \mathbb{P}_i(X_i)$$

$\mathbb{P}_i$ is a distribution over $[s] = \{1,2,\dots,s\}$

Random sample $X = (X_1, X_2, \dots, X_n) \sim \mathbb{P}$

$X \in [s]^n$: n-dimensional random vector

$X_i \in [s]$: independent sample from $\mathbb{P}_i$

$\mathbb{P}$ can be described by $\{\, \mathbb{P}_i : [s] \to [0,1] \mid 1 \le i \le n \,\}$

description size
sn

state space size
s^n

Compute TV distance between product distributions

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

- **Input:** distributions $\{\mathbb{P}_i, \mathbb{Q}_i | 1 \leq i \leq n\}$ specifying $\mathbb{P}$ and $\mathbb{Q}$ over $[s]^n$
- **Output:** the total variation distance between $\mathbb{P}$ and $\mathbb{Q}$

Theorem [BGMMPV22]: the problem is **#P-complete** even for Boolean case $(s = 2)$

FPTAS (Full Poly-time Approximation Scheme)

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

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

FPRAS (Full Poly-time Randomised Approximation Scheme)

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

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

Previous results

Theorem [BGMMPV22] ***FPTAS/FPRAS*** exists for product distributions $\mathbb{P}, \mathbb{Q}$ such that

- $\mathbb{P}$ and $\mathbb{Q}$ are *Boolean* distributions ($s = 2$)
- $\mathbb{Q}$ has *constant number* of distinct marginals (e.g. uniform distribution over $\{0,1\}^n$)

Theorem [BGMMPV22] ***FPRAS*** exists for product distributions $\mathbb{P}, \mathbb{Q}$ such that

- $\mathbb{P}$ and $\mathbb{Q}$ are *Boolean* distributions ($s = 2$)
- $\forall i \in [n], \underbrace{\mathbb{P}_i(1) \geq \mathbb{Q}_i(1)}_{\text{break symmetry}}$ and $\underbrace{\mathbb{P}_i(1) \geq 1/2}_{\text{marginal lower bound}}$

Open problem [BGMMPV22]**:**

Do FPTAS/FPRAS exist for ***general*** product distributions?

Our results [F., Guo, Jerrum, Wang 2023] [F., Liu, Liu 2023] **:**

FPTAS/FPRAS exist for ***general*** product distributions

Product distributions $\mathbb{P}, \mathbb{Q}$ over $[s]^n$ and error bound $0 < \epsilon < 1$

- FPTAS running time: $\tilde{O}\left(\frac{sn^2}{\epsilon}\log\frac{1}{d_{TV}(\mathbb{P},\mathbb{Q})}\right)$
- FPRAS running time : $\tilde{O}\left(\frac{sn^2}{\epsilon^2}\right)$

Extension: Markov chains [F., Liu, Liu 2023]

- distributions π_1, π_2 and transition Matrices M_1, M_2 over state space $[s]$
- approximate $d_{TV}((X_k)_{k=1}^n, (Y_k)_{k=1}^n)$ such that
  - $X_1 \sim \pi_1$ and $X_k \sim M_1(X_{k-1},\cdot)$ / $Y_1 \sim \pi_2$ and $Y_k \sim M_2(Y_{k-1},\cdot)$

FPTAS exists for TV-distance between Markov chains

A natural estimator

Total variation (TV) distance between $\mathbb{P}$ and $\mathbb{Q}$ over state space Ω

$$d_{TV}(\mathbb{P},\mathbb{Q}) = \frac{1}{2}\sum_{x\in\Omega} |\mathbb{P}(x) - \mathbb{Q}(x)| = \sum_{x\in\Omega:\mathbb{Q}(X)>\mathbb{P}(X)} |\mathbb{Q}(x) - \mathbb{P}(x)| = \sum_{x\in\Omega:\mathbb{Q}(X)>\mathbb{P}(X)} \mathbb{Q}(X)\left(1 - \frac{\mathbb{P}(X)}{\mathbb{Q}(X)}\right)$$

Ratio $R \sim \mathbb{R} = (\mathbb{P}||\mathbb{Q})$

$$R = \frac{\mathbb{P}(X)}{\mathbb{Q}(X)}, \qquad \text{where } X \sim \mathbb{Q}$$

$$d_{TV}(\mathbb{P},\mathbb{Q}) = \mathbb{E}[\max(0,1-R)]$$

- sample R independent
- take average of $\max(0,1-R)$

unbiased estimator of $d_{TV}(\mathbb{P},\mathbb{Q})$

- Approximate the TV distance with ***additive*** error $\hat{d} \in d_{TV}(\mathbb{P},\mathbb{Q}) \pm \epsilon$
- ***Relative-error*** approximation requires many samples because

<u>$d_{TV}(\mathbb{P},\mathbb{Q})$ can be exponentially small</u>

TV distance and coupling

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

Coupling Inequality (Coupling Lemma)

$$\forall \text{coupling } (X, Y), \qquad d_{TV}(\mathbb{P}, \mathbb{Q}) \leq \Pr[X \neq Y]$$

$$\exists \text{optimal coupling } (X, Y), \qquad d_{TV}(\mathbb{P}, \mathbb{Q}) = \Pr[X \neq Y]$$

The optimal coupling may ***not*** be unique

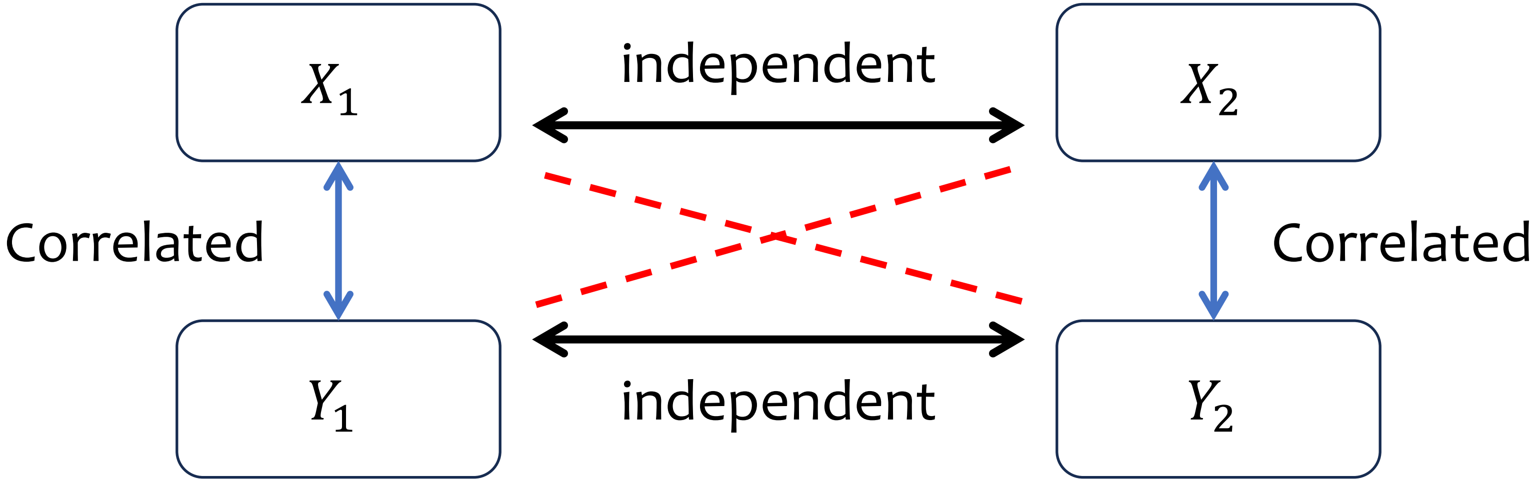
Given two **product distributions** $\mathbb{P}, \mathbb{Q}$ over $[s]^n$,

what is their optimal coupling?

A ***greedy*** coupling $(X, Y) = ((X_1, X_2, \dots, X_n), (Y_1, Y_2, \dots, Y_n))$ of $\mathbb{P}, \mathbb{Q}$

each (X_i, Y_i) is coupled ***optimally*** and ***independently***

- Greedy coupling is **not** an optimal coupling



Optimal coupling can utilise the correlations of (X_1, Y_2) and (Y_1, X_2)

A ***greedy*** coupling $(X,Y)=((X_1,X_2,\dots,X_n),(Y_1,Y_2,\dots,Y_n))$ of $\mathbb{P},\mathbb{Q}$

each (X_i,Y_i) is coupled ***optimally*** and ***independently***

- Greedy coupling is **not** an optimal coupling
- Greedy coupling can **approximate** the optimal coupling

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

Proof.

$$\Pr_{\text{Greedy}}[X \ne Y] \le \sum_{i=1}^{n} \Pr[X_i \ne Y_i] = \sum_{i=1}^{n} d_{TV}(\mathbb{P}_i,\mathbb{Q}_i) \le n d_{TV}(\mathbb{P},\mathbb{Q})$$

local optimal coupling

A ***greedy*** coupling $(X,Y) = ((X_1, X_2, \dots, X_n), (Y_1, Y_2, \dots, Y_n))$ of $\mathbb{P}, \mathbb{Q}$

each (X_i, Y_i) is coupled ***optimally*** and ***independently***

- Greedy coupling is **not** an optimal coupling
- Greedy coupling can **approximate** the optimal coupling

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

- Discrepancy of greedy coupling can be computed ***efficiently***

$$\Pr_{\text{greedy}}[X \ne Y] = 1 - \Pr_{\text{greedy}}[X = Y] = 1 - \prod_{i=1}^{n} \left(1 - d_{TV}(\mathbb{P}_i, \mathbb{Q}_i)\right)$$

- Our ideal: estimate $\frac{\Pr_{\text{opt}}[X \ne Y]}{\Pr_{\text{greedy}}[X \ne Y]} = \frac{d_{TV}(\mathbb{P}, \mathbb{Q})}{\Pr_{\text{greedy}}[X \ne Y]} \ge \frac{1}{n}$

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

- π: the distribution of X in the greedy coupling conditional on $X \neq Y$
$$\forall \sigma \in [s]^n, \qquad \pi(\sigma) = \Pr_{\text{greedy}}[X = \sigma \mid X \neq Y]$$
- f: a function $[s]^n \to \mathbb{R}_{>0}$ such that
$$\forall \sigma \in [s]^n, \qquad f(\sigma) = \frac{\Pr_{\text{opt}}[X = \sigma \wedge X \neq Y]}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]} = \frac{\max\{0, \mathbb{P}(\sigma) - \mathbb{Q}(\sigma)\}}{\Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]}$$
- Estimator: $f(\sigma)$ where $\sigma \sim \pi$

Properties of the estimator

- **Correct expectation**
$$\mathbb{E}_{\sigma\sim\pi}[f(\sigma)] = \frac{\Pr_{\text{opt}}[X \neq Y]}{\Pr_{\text{greedy}}[X \neq Y]} = \frac{d_{TV}(\mathbb{P}, \mathbb{Q})}{\Pr_{\text{greedy}}[X \neq Y]} \geq \frac{1}{n}$$
- **Low variance**
$$\mathrm{Var}_{\sigma\sim\pi}[f(\sigma)] \leq 1$$

$$\forall \sigma \in [s]^V, \qquad \Pr_{\text{opt}}[X = \sigma \wedge X \neq Y] \leq \Pr_{\text{greedy}}[X = \sigma \wedge X \neq Y]$$

⇒

$$\forall \sigma \in [s]^V, \qquad 0 \leq f(\sigma) \leq 1$$

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

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

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

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

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

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

Properties of the estimator

- **Correct expectation**

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

- **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 $O(n)$

$O\left(\frac{n}{\epsilon^2}\right)$ samples

Sampling algorithm for the distribution π:

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

- The greedy coupling is a product distribution
- The condition $X \neq Y$ is not complicated

Algorithm

- Sample $\sigma \in [s]^V$ index by index;
- Conditional on $\sigma_1, \sigma_2, \dots, \sigma_{i-1}$, ***exactly*** compute the marginal of σ_i and sample

$$\Pr[\sigma_1 = c] = \Pr[\, X_1 = c \mid X \neq Y \,] = \frac{\Pr[\, X \neq Y \mid X_1 = c\,] \cdot \Pr[X_1 = c]}{\Pr[X \neq Y]}$$

- ✅ $\Pr[X_1 = c] = P_1(c)$
- ✅ $\Pr[X \neq Y] = 1 - \Pr[X = Y] = 1 - \prod_{i=1}^{n}(1 - d_{TV}(P_i, Q_i))$
- ✅ $\Pr[X \neq Y | X_1 = c] = 1 - \Pr[\, X = Y \mid X_1 = c \,]$

Sampling algorithm for the distribution π:

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

- The greedy coupling is a product distribution
- The condition $X \neq Y$ is not complicated

Algorithm

- Sample $\sigma \in [s]^V$ index by index;
- Conditional on $\sigma_1, \sigma_2, \ldots, \sigma_{i-1}$, ***exactly*** compute the marginal of σ_i and sample

$$\Pr[\sigma_1 = c] = \Pr[X_1 = c \mid X \neq Y] = \frac{\Pr[X \neq Y \mid X_1 = c] \cdot \Pr[X_1 = c]}{\Pr[X \neq Y]}$$

☑ $\Pr[X_1 = c] = P_1(c)$

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

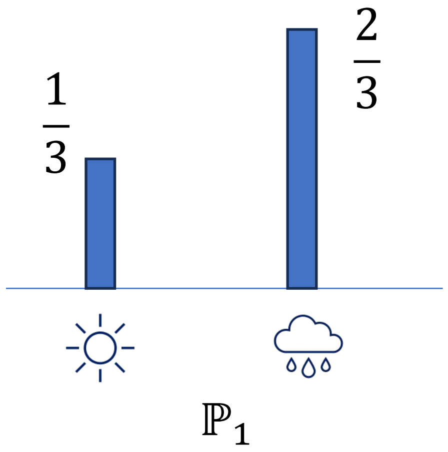
☑ $\Pr[X \neq Y | X_1 = c] = 1 - \Pr[X = Y \mid X_1 = c] = 1 - \Pr[X_1 = Y_1 \mid X_1 = c] \prod_{i=2}^{n} \Pr[X_i = Y_i]$

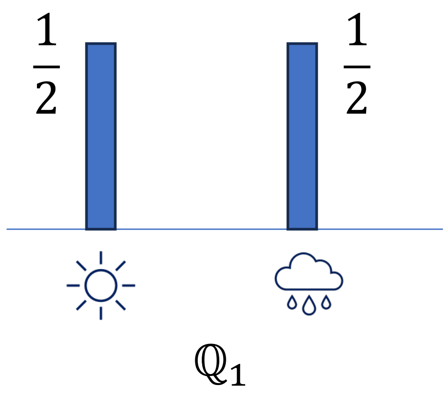
Ratio and deterministic algorithm

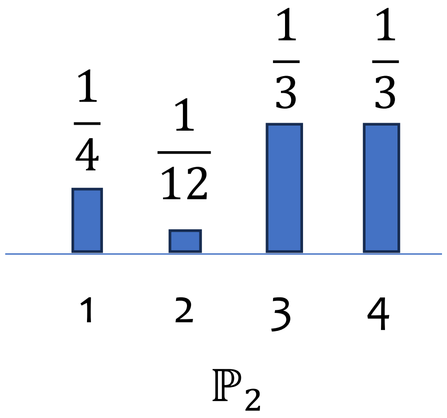
Ratio $R \sim \mathbb{R} = (\mathbb{P}||\mathbb{Q})$

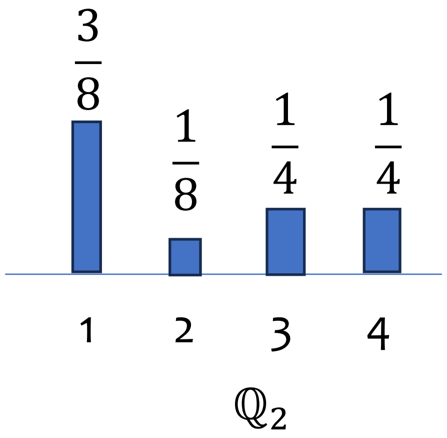
$$R = \frac{\mathbb{P}(X)}{\mathbb{Q}(X)}, \qquad \text{where } X \sim Q$$

$$d_{TV}(\mathbb{P},\mathbb{Q}) = d_{TV}(\mathbb{R}) = \mathbb{E}[\max(0,1-R)]$$



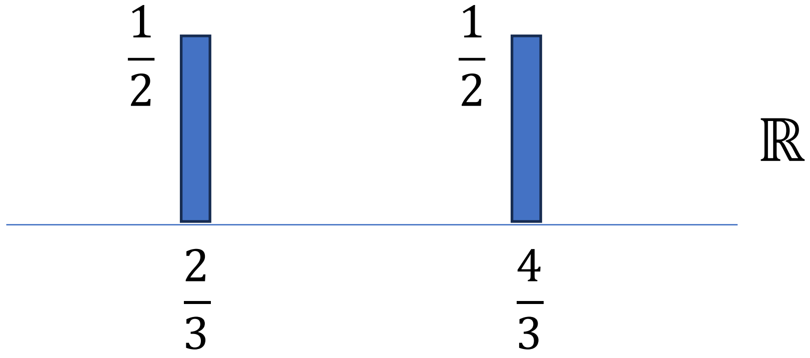






$$\mathbb{R} = (\mathbb{P}_1||\mathbb{Q}_1) = (\mathbb{P}_2||\mathbb{Q}_2)$$

$$d_{TV}(\mathbb{R}) = d_{TV}(\mathbb{P}_1,\mathbb{Q}_1) = d_{TV}(\mathbb{P}_2,\mathbb{Q}_2) = \frac{1}{6}$$



Ratio and deterministic algorithm

Ratio $R \sim \mathbb{R} = (\mathbb{P}||\mathbb{Q})$

$$R = \frac{\mathbb{P}(X)}{\mathbb{Q}(X)}, \qquad \text{where } X \sim Q$$

$$d_{TV}(\mathbb{P}, \mathbb{Q}) = d_{TV}(\mathbb{R}) = \mathbb{E}[\max(0, 1 - R)]$$

$\mathbb{R}$ ***preserves*** $d_{TV}(\mathbb{P}, \mathbb{Q})$, may compress some redundant information

If $\mathbb{R}_1 = (\mathbb{P}_1||\mathbb{Q}_1)$ and $\mathbb{R}_2 = (\mathbb{P}_2||\mathbb{Q}_2)$, then

$$\mathbb{R}_1 \cdot_{\text{ind}} \mathbb{R}_2 = (\mathbb{P}_1 \times \mathbb{P}_2 || \mathbb{Q}_1 \times \mathbb{Q}_2)$$

- $\mathbb{P}_1 \times \mathbb{P}_2$ is the **product distribution** $\mathbb{P}_1$ and $\mathbb{P}_2$
- $\mathbb{R}_1 \cdot_{\text{ind}} \mathbb{R}_2$ is the distribution of the **product of two independent random real number**

$$R_1 R_2 \sim \mathbb{R}_1 \cdot_{\text{ind}} \mathbb{R}_2, \ \text{where } R_1 \sim \mathbb{R}_1 \text{ and } R_2 \sim \mathbb{R}_2 \text{ are ind. samples}$$

A naïve deterministic algorithm

- **Input**: distributions of $\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_n, \mathbb{Q}_1, \mathbb{Q}_2, \ldots, \mathbb{Q}_n$ and an error bound ϵ
- **Output**: an approximation of $d_{TV}(\mathbb{P}, \mathbb{Q})$

- Compute $\mathbb{R}_i \leftarrow (\mathbb{P}_i || \mathbb{Q}_i)$ for all $i \in [n]$

- Compute $\mathbb{R}_{1:1} \leftarrow \mathbb{R}_1$
- Compute $\mathbb{R}_{1:2} \leftarrow \mathbb{R}_{1:1} \cdot_{ind} \mathbb{R}_2$

…

- Compute $\mathbb{R}_{1:i} \leftarrow \mathbb{R}_{1:i-1} \cdot_{ind} \mathbb{R}_i$

…

- Compute distribution $\mathbb{R}_{1:n} \leftarrow \mathbb{R}_{1:n-1} \cdot_{ind} \mathbb{R}_n$
- Return $d_{TV}(\mathbb{R}_{1:n}) = \mathbb{E}_{R \sim \mathbb{R}_{1:n}}[\max(0,1-R)]$

Exact computing of $d_{TV}(\mathbb{P}, \mathbb{Q})$

The support size is ***large***

$$|\text{supp}(\mathbb{R}_{1:i})| = \exp(\Omega(i))$$

Efficiently compute $\widehat{\mathbb{R}}_{\mathbf{1:n}}$ s.t.

$$d_{TV}\left(\widehat{\mathbb{R}}_{1:\boldsymbol{n}}\right) \approx d_{TV}(\mathbb{R}_{1:n})$$

- Compute $\mathbb{R}_i \leftarrow (\mathbb{P}_i || \mathbb{Q}_i)$ for all $i \in [n]$

- $\widehat{\mathbb{R}}_{1:1} \leftarrow \mathbb{R}_1$

- Compute $\mathbb{R}'_{1:2} \leftarrow \widehat{\mathbb{R}}_{1:1} \cdot_{ind} \mathbb{R}_2$
- $\widehat{\mathbb{R}}_{1:2} \leftarrow \textbf{Sparsify}(\mathbb{R}'_{1:2})$

…
- Compute $\mathbb{R}'_{1:i} \leftarrow \widehat{\mathbb{R}}_{1:i-1} \cdot_{ind} \mathbb{R}_i$
- $\widehat{\mathbb{R}}_{1:i} \leftarrow \textbf{Sparsify}(\mathbb{R}'_{1:i})$

…

- Compute $\mathbb{R}'_{1:n} = \widehat{\mathbb{R}}_{1:n-1} \cdot_{ind} \mathbb{R}_n$
- $\widehat{\mathbb{R}}_{1:n} \leftarrow \textbf{Sparsify}(\mathbb{R}'_{1:n})$

- Return $d_{TV}\left(\widehat{\mathbb{R}}_{1:n}\right) = \mathbb{E}_{R\sim\widehat{\mathbb{R}}_{1:n}}[\max(0,1-R)]$

$$\widehat{\mathbb{R}} \leftarrow \text{Sparsify}(\mathbb{R}')$$

- The support size of $\widehat{\mathbb{R}}$ is small
- $\widehat{\mathbb{R}}$ and $\mathbb{R}'$ is close with respect to a ***metric*** $\Delta(\cdot,\cdot)$

- If $\mathbb{R}_1 \approx \mathbb{R}_2$, then $d_{TV}(\mathbb{R}_1) \approx d_{TV}(\mathbb{R}_2)$

$$|d_{TV}(\mathbb{R}_1) - d_{TV}(\mathbb{R}_2)| \leq \Delta(\mathbb{R}_1, \mathbb{R}_2)$$

- If $\mathbb{R}_1 \approx \mathbb{R}_2$ and $\mathbb{R}_3 \approx \mathbb{R}_4$, then $\mathbb{R}_1 \cdot_{ind} \mathbb{R}_3 \approx \mathbb{R}_2 \cdot_{ind} \mathbb{R}_4$

$$\Delta(\mathbb{R}_1 \cdot_{ind} \mathbb{R}_3, \mathbb{R}_2 \cdot_{ind} \mathbb{R}_4) \leq \Delta(\mathbb{R}_1, \mathbb{R}_2) + \Delta(\mathbb{R}_3, \mathbb{R}_4)$$

$$\Delta(\mathbb{R}_1, \mathbb{R}_2) = \min\left\{ d_{TV}(\mathbb{P}_1, \mathbb{P}_2) + d_{TV}(\mathbb{Q}_1, \mathbb{Q}_2) \;\middle|\; \substack{\mathbb{R}_1 = (\mathbb{P}_1 || \mathbb{Q}_1) \\ \mathbb{R}_2 = (\mathbb{P}_2 || \mathbb{Q}_2)} \right\}$$

minimum total variation distance

- Compute $\mathbb{R}_i \leftarrow (\mathbb{P}_i || \mathbb{Q}_i)$ for all $i \in [n]$

$$\mathbb{R}_{i:n} = \mathbb{R}_i \cdot \mathbb{R}_{i+1} \cdot \ldots \cdot \mathbb{R}_n$$

- $\widehat{\mathbb{R}}_{1:1} \leftarrow \mathbb{R}_1$

$$\Delta\left(\widehat{\mathbb{R}}_{1:1}, \mathbb{R}_1\right) = 0 \leq \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

- Compute $\mathbb{R}'_{1:2} \leftarrow \widehat{\mathbb{R}}_{1:1} \cdot_{ind} \mathbb{R}_2$
- $\widehat{\mathbb{R}}_{1:2} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:2})$

$$\Delta\left(\widehat{\mathbb{R}}_{1:2}, \widehat{\mathbb{R}}_{1:1} \cdot \mathbb{R}_2\right) \leq \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

…

- Compute $\mathbb{R}'_{1:i} \leftarrow \widehat{\mathbb{R}}_{i-1} \cdot_{ind} \mathbb{R}_i$
- $\widehat{\mathbb{R}}_{1:i} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:i})$

$$\Delta\left(\widehat{\mathbb{R}}_{1:i}, \widehat{\mathbb{R}}_{1:i-1} \cdot \mathbb{R}_i\right) \leq \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

…

- Compute $\mathbb{R}'_{1:n} = \widehat{\mathbb{R}}_{n-1} \cdot_{ind} \mathbb{R}_n$
- $\widehat{\mathbb{R}}_{1:n} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:n})$

$$\Delta\left(\widehat{\mathbb{R}}_{1:n}, \widehat{\mathbb{R}}_{1:n-1} \cdot \mathbb{R}_n\right) \leq \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

- Return $d_{TV}\left(\widehat{\mathbb{R}}_{1:n}\right) = \mathbb{E}_{R \sim \widehat{\mathbb{R}}_{1:n}}[\max(0, 1 - R)]$

- Compute $\mathbb{R}_i \leftarrow (\mathbb{P}_i || \mathbb{Q}_i)$ for all $i \in [n]$

- $\widehat{\mathbb{R}}_{1:1} \leftarrow \mathbb{R}_1$

- Compute $\mathbb{R}'_{1:2} \leftarrow \widehat{\mathbb{R}}_{1:1} \cdot_{ind} \mathbb{R}_2$
- $\widehat{\mathbb{R}}_{1:2} \leftarrow \textbf{Sparsify}(\mathbb{R}'_{1:2})$

…

- Compute $\mathbb{R}'_{1:i} \leftarrow \widehat{\mathbb{R}}_{i-1} \cdot_{ind} \mathbb{R}_i$
- $\widehat{\mathbb{R}}_{1:i} \leftarrow \textbf{Sparsify}(\mathbb{R}'_{1:i})$

…

$\Delta\left(\widehat{\mathbb{R}}_{1:i} \cdot \mathbb{R}_{i+1:n}, \widehat{\mathbb{R}}_{1:i-1} \cdot \mathbb{R}_i \cdot \mathbb{R}_{i+1:n}\right)$

- Compute $\mathbb{R}'_{1:n} = \widehat{\mathbb{R}}_{n-1} \cdot_{ind} \mathbb{R}_n$
- $\widehat{\mathbb{R}}_{1:n} \leftarrow \textbf{Sparsify}(\mathbb{R}'_{1:n})$

- Return $d_{TV}\left(\widehat{\mathbb{R}}_{1:n}\right) = \mathbb{E}_{R\sim\widehat{\mathbb{R}}_{1:n}}[\max(0,1-R)]$

- Compute $\mathbb{R}_i \leftarrow (\mathbb{P}_i || \mathbb{Q}_i)$ for all $i \in [n]$

- $\widehat{\mathbb{R}}_{1:1} \leftarrow \mathbb{R}_1$

- Compute $\mathbb{R}'_{1:2} \leftarrow \widehat{\mathbb{R}}_{1:1} \cdot_{ind} \mathbb{R}_2$
- $\widehat{\mathbb{R}}_{1:2} \leftarrow \textbf{Sparsify}(\mathbb{R}'_{1:2})$

…

- Compute $\mathbb{R}'_{1:i} \leftarrow \widehat{\mathbb{R}}_{i-1} \cdot_{ind} \mathbb{R}_i$
- $\widehat{\mathbb{R}}_{1:i} \leftarrow \textbf{Sparsify}(\mathbb{R}'_{1:i})$

…

$$\Delta\left(\widehat{\mathbb{R}}_{1:i} \cdot \mathbb{R}_{i+1:n}, \widehat{\mathbb{R}}_{1:i-1} \cdot \mathbb{R}_i \cdot \mathbb{R}_{i+1:n}\right)$$
$$\leq \Delta\left(\widehat{\mathbb{R}}_{1:i}, \widehat{\mathbb{R}}_{1:i-1} \cdot \mathbb{R}_i\right) + \Delta\left(\mathbb{R}_{i+1:n}, \mathbb{R}_{i+1,n}\right)$$
$$\leq 0 + \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

- Compute $\mathbb{R}'_{1:n} = \widehat{\mathbb{R}}_{n-1} \cdot_{ind} \mathbb{R}_n$
- $\widehat{\mathbb{R}}_{1:n} \leftarrow \textbf{Sparsify}(\mathbb{R}'_{1:n})$

- Return $d_{TV}\left(\widehat{\mathbb{R}}_{1:n}\right) = \mathbb{E}_{R \sim \widehat{\mathbb{R}}_{1:n}}[\max(0,1 - R)]$

- Compute $\mathbb{R}_i \leftarrow (\mathbb{P}_i || \mathbb{Q}_i)$ for all $i \in [n]$

- $\widehat{\mathbb{R}}_{1:1} \leftarrow \mathbb{R}_1$

$$\Delta\left(\widehat{\mathbb{R}}_{1:1} \cdot \mathbb{R}_{2:n}, \mathbb{R}_{1:n}\right) \le \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

- Compute $\mathbb{R}'_{1:2} \leftarrow \widehat{\mathbb{R}}_{1:1} \cdot_{ind} \mathbb{R}_2$
- $\widehat{\mathbb{R}}_{1:2} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:2})$

$$\Delta\left(\widehat{\mathbb{R}}_{1:2} \cdot \mathbb{R}_{3:n}, \widehat{\mathbb{R}}_{1:1} \cdot \mathbb{R}_{2:n}\right) \le \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

…

- Compute $\mathbb{R}'_{1:i} \leftarrow \widehat{\mathbb{R}}_{i-1} \cdot_{ind} \mathbb{R}_i$
- $\widehat{\mathbb{R}}_{1:i} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:i})$

$$\Delta\left(\widehat{\mathbb{R}}_{1:i} \cdot \mathbb{R}_{i+1:n}, \widehat{\mathbb{R}}_{1:i-1} \cdot \mathbb{R}_{i:n}\right) \le \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

…

- Compute $\mathbb{R}'_{1:n} = \widehat{\mathbb{R}}_{n-1} \cdot_{ind} \mathbb{R}_n$
- $\widehat{\mathbb{R}}_{1:n} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:n})$

$$\Delta\left(\widehat{\mathbb{R}}_{1:n}, \widehat{\mathbb{R}}_{1:n-1} \cdot \mathbb{R}_n\right) \le \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

- Return $d_{TV}\left(\widehat{\mathbb{R}}_{1:n}\right) = \mathbb{E}_{R \sim \widehat{\mathbb{R}}_{1:n}}[\max(0,1-R)]$

- Compute $\mathbb{R}_i \leftarrow (\mathbb{P}_i || \mathbb{Q}_i)$ for all $i \in [n]$

- $\widehat{\mathbb{R}}_{1:1} \leftarrow \mathbb{R}_1$

- Compute $\mathbb{R}'_{1:2} \leftarrow \widehat{\mathbb{R}}_{1:1} \cdot_{ind} \mathbb{R}_2$
- $\widehat{\mathbb{R}}_{1:2} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:2})$

…

- Compute $\mathbb{R}'_{1:i} \leftarrow \widehat{\mathbb{R}}_{i-1} \cdot_{ind} \mathbb{R}_i$
- $\widehat{\mathbb{R}}_{1:i} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:i})$

…

- Compute $\mathbb{R}'_{1:n} = \widehat{\mathbb{R}}_{n-1} \cdot_{ind} \mathbb{R}_n$
- $\widehat{\mathbb{R}}_{1:n} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:n})$

- Return $d_{TV}\left(\widehat{\mathbb{R}}_{1:n}\right) = \mathbb{E}_{R \sim \widehat{\mathbb{R}}_{1:n}}[\max(0, 1 - R)]$

$$\Delta\left(\widehat{\mathbb{R}}_{1:1} \cdot \mathbb{R}_{2:n}, \mathbb{R}_{1:n}\right) \leq \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

$$\Delta\left(\widehat{\mathbb{R}}_{1:2} \cdot \mathbb{R}_{3:n}, \widehat{\mathbb{R}}_{1:1} \cdot \mathbb{R}_{2:n}\right) \leq \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

$$\Delta\left(\widehat{\mathbb{R}}_{1:i} \cdot \mathbb{R}_{i+1:n}, \widehat{\mathbb{R}}_{1:i-1} \cdot \mathbb{R}_{i:n}\right) \leq \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

$$\Delta\left(\widehat{\mathbb{R}}_{1:n}, \widehat{\mathbb{R}}_{1:n-1} \cdot \mathbb{R}_n\right) \leq \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

- Compute $\mathbb{R}_i \leftarrow (\mathbb{P}_i || \mathbb{Q}_i)$ for all $i \in [n]$

- $\widehat{\mathbb{R}}_{1:1} \leftarrow \mathbb{R}_1$

- Compute $\mathbb{R}'_{1:2} \leftarrow \widehat{\mathbb{R}}_{1:1} \cdot_{ind} \mathbb{R}_2$
- $\widehat{\mathbb{R}}_{1:2} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:2})$

…

- Compute $\mathbb{R}'_{1:i} \leftarrow \widehat{\mathbb{R}}_{i-1} \cdot_{ind} \mathbb{R}_i$
- $\widehat{\mathbb{R}}_{1:i} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:i})$

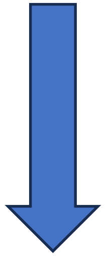
…

- Compute $\mathbb{R}'_{1:n} = \widehat{\mathbb{R}}_{n-1} \cdot_{ind} \mathbb{R}_n$
- $\widehat{\mathbb{R}}_{1:n} \leftarrow$ **Sparsify**$(\mathbb{R}'_{1:n})$

- Return $d_{TV}\left(\widehat{\mathbb{R}}_{1:n}\right) = \mathbb{E}_{R\sim\widehat{\mathbb{R}}_{1:n}}[\max(0,1-R)]$

By **triangle-inequality** of metric

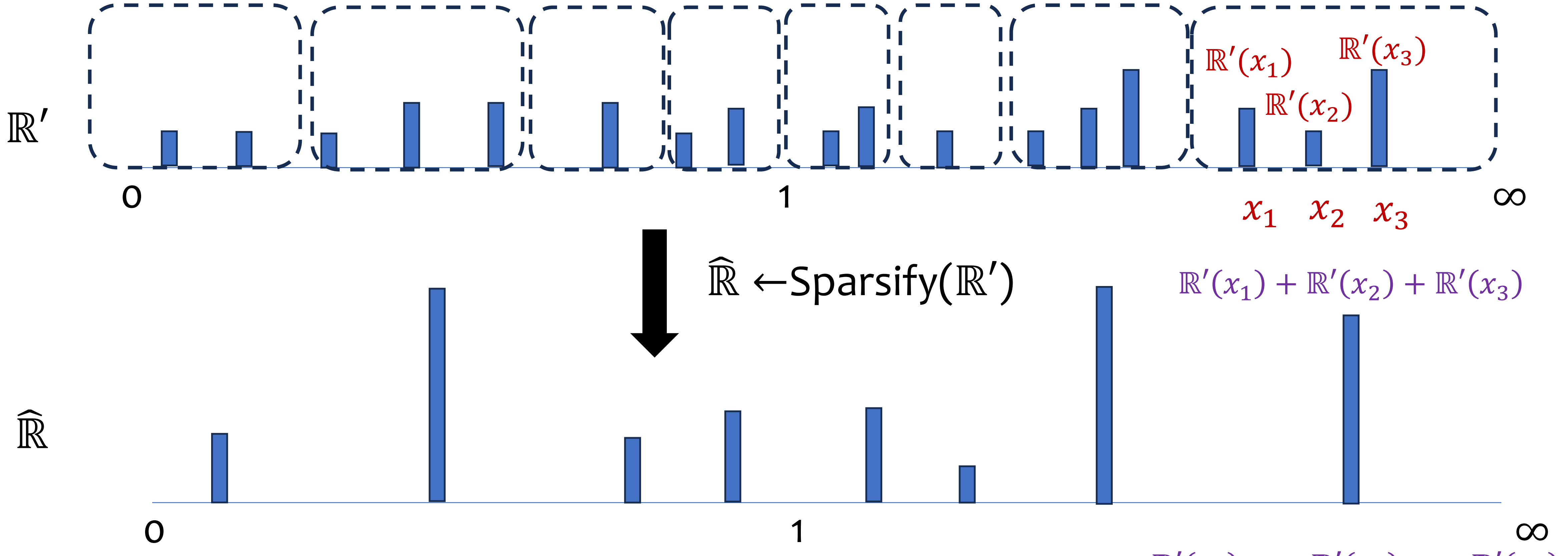
$$\Delta\left(\mathbb{R}_{1:n}, \widehat{\mathbb{R}}_{1:n}\right) \leq \epsilon d_{TV}(\mathbb{P}, \mathbb{Q})$$



$$\mathbb{R}_{1:n} = (\mathbb{P}||\mathbb{Q})$$
$$d_{TV}(\mathbb{R}_{1:n}) = d_{TV}(\mathbb{P}, \mathbb{Q})$$

$$\left|d_{TV}\left(\widehat{\mathbb{R}}_{1:n}\right) - d_{TV}(\mathbb{P}, \mathbb{Q})\right| \leq \epsilon d_{TV}(\mathbb{P}, \mathbb{Q})$$

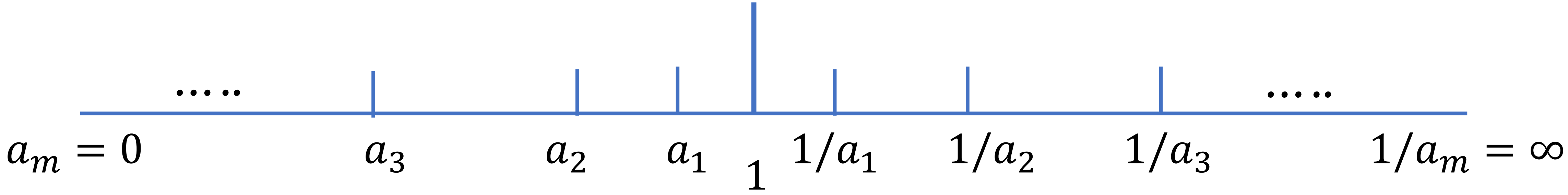
The Sparsify subroutine



$$\frac{x_1\mathbb{R}'(x_1) + x_2\mathbb{R}'(x_2) + x_3\mathbb{R}'(x_3)}{\mathbb{R}'(x_1) + \mathbb{R}'(x_2) + \mathbb{R}'(x_3)}$$

$$= \mathbb{E}_{R' \sim \mathbb{R}'}[R' \mid R' \in I]$$

- Partition $[0, \infty)$ into a set of intervals $\mathcal{I}$
- For each interval $I \in \mathcal{I}$, merge all elements in I into one element



Partition of [0,1]

$$\left\{ \begin{array}{c} [a_1, a_0 = 1) \\ [a_2, a_1) \\ [a_3, a_2) \\ \dots \\ [a_m = 0, a_{m-1}) \end{array} \right.$$

- The first interval is small

$$1 - a_1 \le \delta_s$$

- The length of $[a_i, a_{i-1}]$ is small w.r.t. $1 - a_{i-1}$

$$\forall i > 1, \qquad |a_i - a_{i-1}| \le \epsilon_s \cdot |1 - a_{i-1}|$$

Partition of $(1, \infty)$

$$\left\{ \begin{array}{c} [a_0 = 1, 1/a_1) \\ [1/a_1, 1/a_2) \\ [1/a_2, 1/a_3) \\ \dots \\ [1/a_{m-1}, 1/a_m = \infty) \end{array} \right.$$

$$m = O\left(\frac{1}{\epsilon_s} \log \frac{1}{\delta_s}\right)$$

Error of Sparsification [F., Liu, Liu 2023]

$\mathbb{R} \leftarrow$Sparsify($\mathbb{R}'$) $\Rightarrow$ $\Delta(\mathbb{R}, \mathbb{R}') \leq \epsilon_s d_{TV}(\mathbb{R}') + \delta_s$

- δ_s: ***absolute error*** from merging $[a_1, 1]$ and $\left(1, \frac{1}{a_1}\right)$
- ϵ_s: ***relative error*** from merging other intervals

$$\frac{\epsilon d_{TV}(\mathbb{P}, \mathbb{Q})}{2n^2} \leq \delta_s \leq \frac{\epsilon d_{TV}(\mathbb{P}, \mathbb{Q})}{2n}$$

$$\epsilon_s = \frac{\epsilon}{2n}$$

Merge error in every iteration

$$\Delta\left(\widehat{\mathbb{R}}_{1:i}, \widehat{\mathbb{R}}_{1:i-1} \cdot \mathbb{R}_i\right)$$

$$\leq \frac{\epsilon}{2n} d_{TV}\left(\widehat{\mathbb{R}}_{1:i-1} \cdot \mathbb{R}_i\right) + \frac{d_{TV}(\mathbb{P}, \mathbb{Q})}{2n}$$

$$\leq \frac{\epsilon}{n} d_{TV}(\mathbb{P}, \mathbb{Q})$$

$$m = O\left(\frac{n}{\epsilon} \log \frac{1}{d_{TV}(\mathbb{P}, \mathbb{Q})}\right)$$

Summary

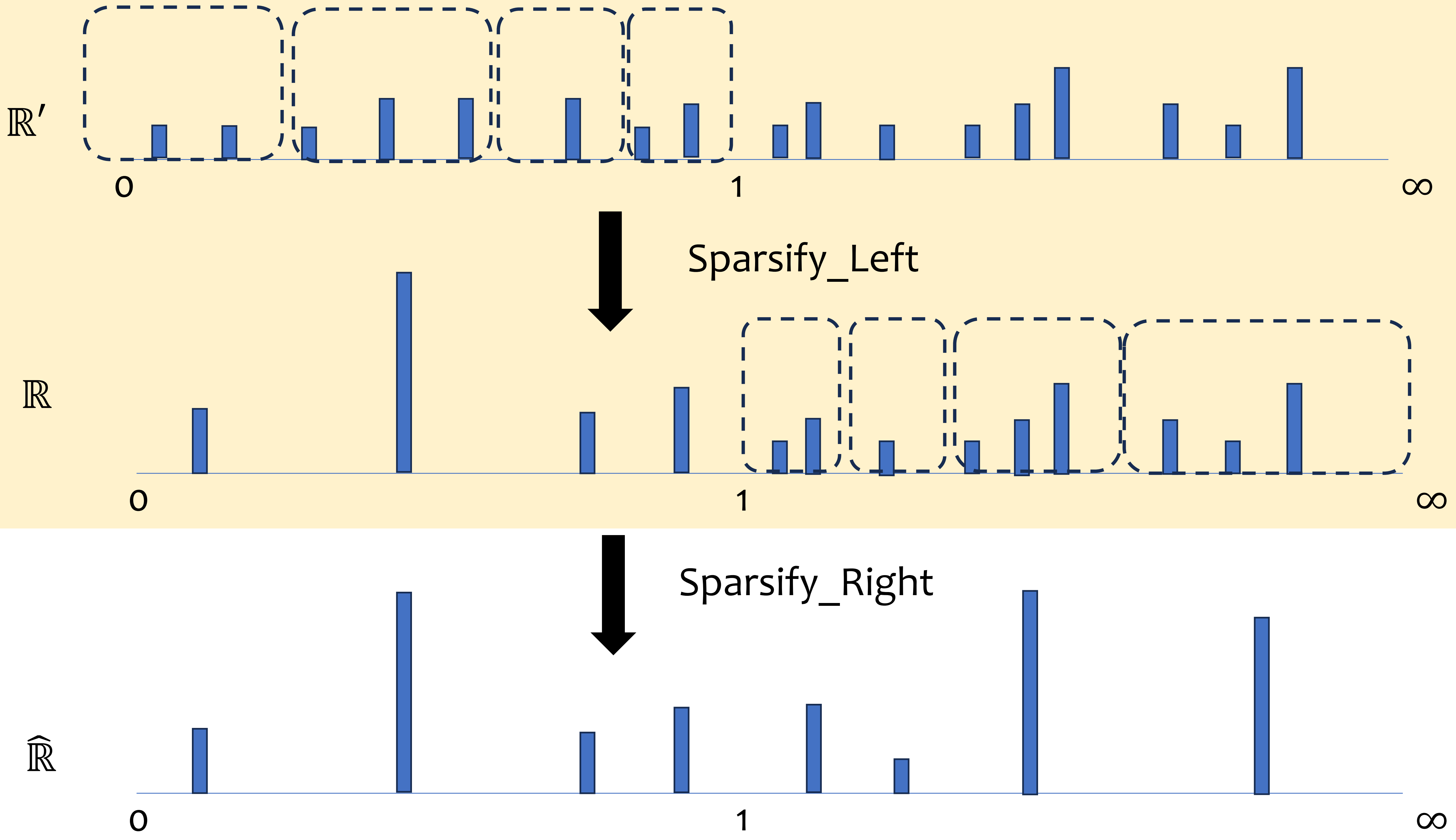
- **Problem:** Compute the TV distance between two ***product distributions***
- **Algorithms:** FPTAS and FPRAS
- **Extension:** TV distance between two ***Markov chains***

Open problems

- ***Better running time*** of FPTAS: remove $\log \frac{1}{d_{TV}(\mathbb{P},\mathbb{Q})}$ in $\tilde{O}(n^2 \log \frac{1}{d_{TV}(\mathbb{P},\mathbb{Q})})$?
- Algorithm/complexity for approximating TV distance of ***general models***
  - Graphical models
  - Hidden Markov chains
- Relation between ***approximating TV distance*** and ***sampling/counting***

Appendix

Analysis of the sparsification error



ℝ′
0
1
∞
Sparsify_Left
ℝ
0
1
∞
Sparsify_Right
ℝ̂
0
1
∞

$\mathbb{R} \leftarrow$ **Sparsify_Left**$(\mathbb{R}')$ merge intervals $\mathcal{I} = \{[a_1, a_0), [a_2, a_1), \ldots, [a_m, a_{m-1})\}$

$\mathbb{R}' = (\mathbb{P}' || \mathbb{Q}')$, where $\mathbb{Q}' = \mathbb{R}'$ and $\mathbb{P}'(r) = r\mathbb{Q}'(r)$

$$\mathbb{P}(r) = \begin{cases} \mathbb{P}'(r) & \text{if } r > 1 \\ \dfrac{\mathbb{Q}'(r)}{\mathbb{Q}'(I)} \mathbb{P}'(I) & \text{if } r \in I \in \mathcal{I} \end{cases}$$

$\Rightarrow$

$\forall r \in I \in \mathcal{I}$
$$\frac{\mathbb{P}(r)}{\mathbb{Q}'(r)} = \frac{\mathbb{P}(I)}{\mathbb{Q}(I)}$$

$$\mathbb{Q}'(I) = \sum_{r \in I} \mathbb{Q}'(r)$$

All the ratios for $r \in I$ are the same
$\approx$ Merge

$\mathbb{R} \leftarrow$ **Sparsify_Left**$(\mathbb{R}')$ merge intervals $\mathcal{I} = \{[a_1, a_0), [a_2, a_1), \ldots, [a_m, a_{m-1})\}$

$\mathbb{R}' = (\mathbb{P}' || \mathbb{Q}')$, where $\mathbb{Q}' = \mathbb{R}'$ and $\mathbb{P}'(r) = r\mathbb{Q}'(r)$

$$\mathbb{P}(r) = \begin{cases} \mathbb{P}'(r) & \text{if } r > 1 \\ \dfrac{\mathbb{Q}'(r)}{\mathbb{Q}'(I)} \mathbb{P}'(I) & \text{if } r \in I \in \mathcal{I} \end{cases}$$

$$\mathbb{R} = (\mathbb{P} || \mathbb{Q}')$$

$$\Delta(\mathbb{R}', \mathbb{R}) \leq d_{TV}(\mathbb{P}', \mathbb{P}) + d_{TV}(\mathbb{Q}', \mathbb{Q}') = \frac{1}{2} \sum_{I \in \mathcal{I}} \sum_{r \in I} |\mathbb{P}'(r) - \mathbb{P}(r)|$$

$\mathbb{R} \leftarrow$ **Sparsify_Left**$(\mathbb{R}')$ merge intervals $L = \{[a_1, a_0), (a_2, a_1), \dots, (a_m, a_{m-1})\}$

$\mathbb{R}' = (\mathbb{P}'||\mathbb{Q}')$, where $\mathbb{Q}' = \mathbb{R}'$ and $\mathbb{P}'(r) = r\mathbb{Q}'(r)$

$$\mathbb{P}(r) = \begin{cases} \mathbb{P}'(r) & \text{if } r > 1 \\ \frac{\mathbb{Q}'(r)}{\mathbb{Q}'(I)}\mathbb{P}'(I) & \text{if } r \in I \in \mathcal{J} \end{cases}$$

$\Rightarrow$

$$\mathbb{R} = (\mathbb{P}||\mathbb{Q}')$$

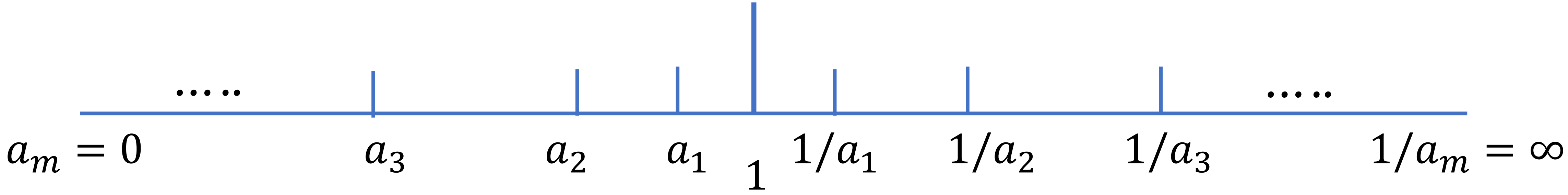
$$\Delta(\mathbb{R}', \mathbb{R}) \le \frac{1}{2}\sum_{I \in L}\sum_{r \in I}\mathbb{Q}'(r)\left|r - \frac{\mathbb{P}'(I)}{\mathbb{Q}'(I)}\right|$$

At most the length of interval $\le |a_i - a_{i-1}|$, where $I = [a_i, a_{i-1})$

vs

$$d_{TV}(\mathbb{R}') = \frac{1}{2}\sum_{I \in L \cup R}\sum_{r}\mathbb{Q}'(r)|r - 1|$$

At least the distance between 1 and right point $\ge 1 - a_{i-1}$



Partition of $[0,1]$

$[a_1, a_0 = 1)$
$[a_2, a_1)$
$[a_3, a_2)$
…
$[a_m, a_{m-1})$

Partition of $(1, \infty)$

$[a_0 = 1, 1/a_1)$
$[1/a_1, 1/a_2)$
$[1/a_2, 1/a_3)$
…
$[1/a_{m-1}, 1/a_m)$

- The first interval is small

$$1 - a_1 \leq \delta_s$$

- The length of $[a_i, a_{i-1}]$ is small w.r.t. $1 - a_{i-1}$

$$\forall i > 1, \qquad |a_i - a_{i-1}| \leq \epsilon_s \cdot |1 - a_{i-1}|$$

$$m = O\left(\frac{1}{\epsilon_s} \log \frac{1}{\delta_s}\right)$$

$\mathbb{R} \leftarrow$ **Sparsify_Left**$(\mathbb{R}')$ merge intervals $L = \{[a_1, a_0), (a_2, a_1), \dots, (a_m, a_{m-1})\}$

$$\mathbb{R}' = (\mathbb{P}' || \mathbb{Q}'), \text{ where } \mathbb{Q}' = \mathbb{R}' \text{ and } \mathbb{P}'(r) = r\mathbb{Q}'(r)$$

$$\mathbb{P}(r) = \begin{cases} \mathbb{P}'(r) & \text{if } r > 1 \\ \dfrac{\mathbb{Q}'(r)}{\mathbb{Q}'(I)} \mathbb{P}'(I) & \text{if } r \in I \in \mathcal{I} \end{cases}$$

$$\mathbb{R} = (\mathbb{P} || \mathbb{Q}')$$

$$\Delta(\mathbb{R}', \mathbb{R}) \leq \frac{1}{2}(\epsilon_s d_{TV}(\mathbb{R}') + \delta_s)$$

Error from merging other intervals

Error from merging the first interval