Anuran Makur
3 May 2017

Logarithmic Sobolev Inequalities in Discrete Product Spaces: (Marton, 2015)

Let $\mathcal{X}$ be a finite set, and $\mathcal{X}^n$ be the set of n-sequences from $\mathcal{X}$.
Let $\mathcal{P}(\mathcal{X}^n)$ be the probability simplex on $\mathcal{X}^n$.

Def: (W_2-distance) Suppose $X^n \sim P_{X^n} \in \mathcal{P}(\mathcal{X}^n)$ and $Y^n \sim P_{Y^n} \in \mathcal{P}(\mathcal{X}^n)$. Then, we define:

$$W_2(P_{X^n}, P_{Y^n}) \triangleq \min_{P_{X^n,Y^n}} \sqrt{\sum_{i=1}^{n} P_{X^n,Y^n}(X_i \neq Y_i)^2}$$

where the minimum is over all couplings of P_{X^n} and P_{Y^n}.

Remark:
- This is not a 2-Wasserstein distance wrt Hamming metric.
- Minimum is achieved.
- W_2 is a metric on $\mathcal{P}(\mathcal{X}^n)$.

Theorem 1: (W_2^2 tensorization $\Rightarrow$ KL tensorization)
Fix a distribution $Q_{X^n} \in \mathcal{P}(\mathcal{X}^n)$, and let $\alpha \triangleq \min_{\substack{1 \leq i \leq n \\ x^n \in \mathcal{X}^n : Q_{X^n}(x^n) > 0}} Q_{X_i|X_{i^c}}(x_i|x_{i^c})$, where for any string x^n and any set $S \subseteq [n] \triangleq \{1,\dots,n\}$, $x_S \triangleq \{x_i : i \in S\}$, and $x_i = x_{\{i\}}$ and $x_{i^c} = x_{\{i\}^c}$.
Fix a distribution $P_{Y^n} \in \mathcal{P}(\mathcal{X}^n)$ such that $P_{Y^n} \overset{\text{absolutely continuous}}{\ll} Q_{X^n}$.
Suppose for all $I \subseteq [n]$ and all $y_{I^c} \in \mathcal{X}^{[n]\setminus I}$, we have:

$$W_2^2(P_{Y_I|Y_{I^c}}(\cdot|y_{I^c}), Q_{X_I|X_{I^c}}(\cdot|y_{I^c})) \leq C \sum_{i \in I} \underset{\text{wrt } P_{Y_{i^c}|Y_{I^c} = y_{I^c}}}{\mathbb{E}}\Big[\overbrace{\left\| P_{Y_i|Y_{i^c}}(\cdot|Y_{i^c}) - Q_{X_i|X_{i^c}}(\cdot|Y_{i^c}) \right\|_{TV}^2}^{W_2^2(P_{Y_i|Y_{i^c}}(\cdot|Y_i^c),\, Q_{X_i|X_{i^c}}(\cdot|Y_i^c))} \,\Big|\, \underset{\text{conditional } W_2^2\text{- distance}}{Y_{I^c} = y_{I^c}}\Big]$$

for a universal constant C. Then, we have:

$$D(P_{Y^n}\|Q_{X^n}) \underset{\text{reverse Pinsker}}{\leq} \frac{4C}{\alpha} \sum_{i=1}^{n} \underset{\text{wrt } P_{Y_i^c}}{\mathbb{E}}\left[\left\| P_{Y_i|Y_{i^c}}(\cdot|Y_{i^c}) - Q_{X_i|X_{i^c}}(\cdot|Y_{i^c}) \right\|_{TV}^2\right] \underset{\text{Pinsker}}{\leq} \frac{2C}{\alpha} \sum_{i=1}^{n} \underset{\text{conditional KL divergence}}{D(P_{Y_i|Y_{i^c}} \| Q_{X_i|X_i^c} | P_{Y_i^c})}.$$

Proof: We prove this by induction. Assume the conditions of the theorem hold, and that we have proven it for $n-1$:

$$D\left(P_{Y_{i^c}|Y_i}(\cdot|y_i) \,\|\, Q_{X_{i^c}|X_i}(\cdot|y_i)\right) \leq \frac{4C}{\alpha} \sum_{j \neq i} \mathbb{E}\left[\left\| P_{Y_j|Y_{j^c}}(\cdot|Y_{j^c}) - Q_{X_j|X_{j^c}}(\cdot|Y_{j^c}) \right\|_{TV}^2 \,\Big|\, Y_i = y_i\right] \leftarrow \text{inductive hypothesis}$$

for all $i \in [n]$ and all $y_i \in \mathcal{X}$.

First, by averaging chain rules, we get:

$$D(P_{Y^n}\|Q_{X^n}) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} D(P_{Y_i}\|Q_{X_i})}_{①} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} D(P_{Y_{i^c}|Y_i} \| Q_{X_{i^c}|X_i} | P_{Y_i})}_{②}. \quad [*]$$

[continued]

Anuran Makur
3 May 2017

Proof continued:

To bound ②, we use the inductive hypothesis to get:

$$\frac{1}{n}\sum_{i=1}^{n} D(P_{Y_{i^c}|Y_i}\|Q_{X_{i^c}|X_i}|P_{Y_i}) \leq \frac{1}{n}\frac{4C}{\alpha}\sum_{i=1}^{n}\underbrace{\sum_{j\neq i}\mathbb{E}\left[\left\|P_{Y_j|Y_{j^c}}(\cdot|Y_{j^c}) - Q_{X_j|X_{j^c}}(\cdot|Y_{j^c})\right\|_{TV}^2\right]}_{\text{does not depend on } i \Rightarrow (n-1) \text{ copies for each } j}$$

$$= \left(1-\frac{1}{n}\right)\frac{4C}{\alpha}\sum_{j=1}^{n}\mathbb{E}\left[\left\|P_{Y_j|Y_{j^c}}(\cdot|Y_{j^c}) - Q_{X_j|X_{j^c}}(\cdot|Y_{j^c})\right\|_{TV}^2\right]. \quad [\star]$$

To bound ①, notice from reverse Pinsker's inequality $\left(D(P\|Q) \leq \frac{4}{\min_x Q(x)}\|P-Q\|_{TV}^2 \text{ for } Q>0\right)$:

$$D(P_{Y_i}\|Q_{X_i}) \leq \frac{4}{\alpha}\|P_{Y_i} - Q_{X_i}\|_{TV}^2$$

for all $i\in[n]$, where we use the fact that: $\forall x\in\mathcal{X}, \forall i\in[n], Q_{X_i}(x) \geq \alpha$.

$$\left[Q_{X_i}(x) = \sum_{x_{i^c}}\underbrace{Q_{X_i|X_{i^c}}(x|x_{i^c})}_{\geq\alpha \text{ if } Q_{X_i,X_{i^c}}(x,x_{i^c})>0}Q_{X_{i^c}}(x_{i^c})\right]$$

Let π be the optimal W_2-coupling of P_{Y^n}, Q_{X^n}. Then, we have:

$$W_2^2(P_{Y^n},Q_{X^n}) = \sum_{i=1}^{n}\mathbb{P}_\pi(X_i\neq Y_i)^2 \geq \sum_{i=1}^{n}\|P_{Y_i}-Q_{X_i}\|_{TV}^2 \quad [\text{maximal coupling of TV}]$$

which implies using the condition in the theorem with $I=[n]$ that:

$$\frac{1}{n}\sum_{i=1}^{n}D(P_{Y_i}\|Q_{X_i}) \leq \frac{1}{n}\frac{4}{\alpha}\sum_{i=1}^{n}\|P_{Y_i}-Q_{X_i}\|_{TV}^2 \leq \frac{1}{n}\frac{4}{\alpha}W_2^2(P_{Y^n},Q_{X^n})$$

$$\leq \frac{1}{n}\frac{4C}{\alpha}\sum_{i=1}^{n}\mathbb{E}\left[\left\|P_{Y_i|Y_{i^c}}(\cdot|Y_{i^c}) - Q_{X_i|X_{i^c}}(\cdot|Y_{i^c})\right\|_{TV}^2\right]. \quad [\star]$$

Hence, we get:

$$D(P_{Y^n}\|Q_{X^n}) \leq \frac{4C}{\alpha}\sum_{i=1}^{n}\mathbb{E}\left[\left\|P_{Y_i|Y_{i^c}}(\cdot|Y_{i^c}) - Q_{X_i|X_{i^c}}(\cdot|Y_{i^c})\right\|_{TV}^2\right] \quad (\text{using } [\star])$$

$$\leq \frac{2C}{\alpha}\sum_{i=1}^{n}D(P_{Y_i|Y_{i^c}}\|Q_{X_i|X_{i^c}}|P_{Y_{i^c}})$$

where the final inequality follows from Pinsker's inequality:

$$D(P\|Q) \geq 2\|P-Q\|_{TV}^2.$$

∎

Remark: Doing the induction from k-1 to k would require us to use more conditional expectations, and would illustrate the full power of the condition in the theorem.

Idea: W_2-distance related to TV distance, KL distance bounded by TV distance (Pinsker + reverse Pinsker), use this to prove W_2^2-tensorization ⇒ KL tensorization.
$(KL \leq TV^2 \leq W_2^2 \leq TV^2 \leq KL)$

[continued.]

Anuran Makur
3 May 2017

<u>Def:</u> (Gibbs Sampler) For $i \in [n]$, define the <u>Markov chain</u>:

$$T_i(x^n|y^n) \triangleq \mathbb{1}\{x_{i^c} = y_{i^c}\} Q_{X_i|X_{i^c}}(x_i|y_{i^c}) \, , \quad \forall x^n, y^n \in \mathcal{X}^n$$

for a distribution $Q_{X^n} \in \mathcal{P}(\mathcal{X}^n)$. The <u>Gibb sampler</u> for Q_{X^n} is the Markov chain:

$$\underbrace{T}_{\text{row stochastic matrices}} \triangleq \frac{1}{n}\sum_{i=1}^{n} T_i \, .$$

Observe that for any $x^n \in \mathcal{X}^n$:

$$Q_{X^n}T_i(x^n) = \sum_{y^n} Q_{X^n}(y^n) T_i(x^n|y^n) = \sum_{y^n} Q_{X^n}(y^n)\mathbb{1}\{x_{i^c} = y_{i^c}\} Q_{X_i|X_{i^c}}(x_i|y_{i^c})$$

$$= Q_{X_i|X_{i^c}}(x_i|x_{i^c}) \sum_{y_i} Q_{X_i, X_{i^c}}(y_i, x_{i^c}) = Q_{X_i|X_{i^c}}(x_i|x_{i^c}) Q_{X_{i^c}}(x_{i^c})$$

$$= Q_{X^n}(x^n)$$

which means Q_{X^n} is the <u>invariant measure</u> for T_i and T.

The <u>Dirichlet form</u> associated with T is:

$$\mathcal{E}_T(f,f) \triangleq \langle (\underbrace{I}_{\text{identity map}} - T)f, f \rangle_{Q_{X^n}} = \sum_{x^n} \Big(f(x^n) - \underbrace{Tf(x^n)}_{= \sum_{y^n} f(y^n) T(y^n|x^n)}\Big) f(x^n) Q_{X^n}(x^n) = \frac{1}{2}\sum_{x^n, y^n} T(y^n|x^n) Q_{X^n}(x^n)\big(f(y^n) - f(x^n)\big)^2$$

for all $f \in \mathcal{L}^2(\mathcal{X}^n, Q_{X^n})$. The <u>logarithmic Sobolev inequality</u> for T is:

$$D(P_{Y^n}\|Q_{X^n}) \leq c\, \mathcal{E}_T\Bigg(\overbrace{\sqrt{\frac{P_{Y^n}}{Q_{X^n}}}}^{f}, \overbrace{\sqrt{\frac{P_{Y^n}}{Q_{X^n}}}}^{f}\Bigg) \quad \text{for all } P_{Y^n} \in \mathcal{P}(\mathcal{X}^n)$$

where the best constant c is called the <u>log-Sobolev constant</u>.

<u>Remark:</u> In some papers, $\frac{1}{c}$ is the log-Sobolev constant.

<u>Corollary:</u> (Gibbs Sampler contraction)

If $Q_{X^n}, P_{Y^n} \in \mathcal{P}(\mathcal{X}^n)$ satisfy the conditions of Theorem 1, then:

[KL Contraction] ① $D(P_{Y^n}T\|Q_{X^n}) \leq \left(1 - \frac{\alpha}{2nC}\right) D(P_{Y^n}\|Q_{X^n})$,

[Log-Sobolev] ② $D(P_{Y^n}\|Q_{X^n}) \leq \frac{4Cn}{\alpha}\, \mathcal{E}_T\left(\sqrt{\frac{P_{Y^n}}{Q_{X^n}}}, \sqrt{\frac{P_{Y^n}}{Q_{X^n}}}\right)$.

[Hold for <u>all</u> P_{Y^n} when Q_{X^n} satisfies <u>Dobrushin uniqueness condition</u>.]

<u>Proof:</u> ① First observe that:

$$D(P_{Y^n}\|Q_{X^n}) = D(P_{Y_{i^c}}\|Q_{X_{i^c}}) + D(P_{Y_i|Y_{i^c}}\|Q_{X_i|X_{i^c}}|P_{Y_{i^c}}) \quad \text{[Chain rule]}$$

$$= D(P_{Y^n}T_i\|Q_{X^n}) + D(P_{Y_i|Y_{i^c}}\|Q_{X_i|X_{i^c}}|P_{Y_{i^c}})$$

$$= \frac{1}{n}\sum_{i=1}^{n} D(P_{Y^n}T_i\|Q_{X^n}) + \underbrace{\frac{1}{n}\sum_{i=1}^{n} D(P_{Y_i|Y_{i^c}}\|Q_{X_i|X_{i^c}}|P_{Y_{i^c}})}_{\geq \frac{\alpha}{2nC} D(P_{Y^n}\|Q_{X^n})}.$$

$$P_{Y^n}T_i(y^n) = \sum_{x^n} P_{Y^n}(x^n)\mathbb{1}\{y_{i^c} = x_{i^c}\} Q_{X_i|X_{i^c}}(y_i|y_{i^c}) = Q_{X_i|X_{i^c}}(y_i|y_{i^c}) P_{Y_{i^c}}(y_{i^c})$$

$$Q_{X^n}(y^n) = Q_{X_i|X_{i^c}}(y_i|y_{i^c}) Q_{X_{i^c}}(y_{i^c})$$

By <u>convexity</u> of KL divergence:

$$D(P_{Y^n}T\|Q_{X^n}) \leq \frac{1}{n}\sum_{i=1}^{n} D(P_{Y^n}T_i\|Q_{X^n}) \leq D(P_{Y^n}\|Q_{X^n}) - \frac{\alpha}{2nC} D(P_{Y^n}\|Q_{X^n})$$

using the KL tensorization result of Theorem 1.

[continued.]

Anuran Makur
3 May 2017

Proof continued:

② First observe that:

$$\mathcal{E}_{T_i}\left(\sqrt{\frac{P_{Y^n}}{Q_{X^n}}},\sqrt{\frac{P_{Y^n}}{Q_{X^n}}}\right) = \left\langle (I-T_i)\sqrt{\frac{P_{Y^n}}{Q_{X^n}}},\sqrt{\frac{P_{Y^n}}{Q_{X^n}}}\right\rangle_{Q_{X^n}} = \sum_{y^n}\frac{P_{Y^n}(y^n)}{Q_{X^n}(y^n)}Q_{X^n}(y^n) - \sum_{y^n} T_i\sqrt{\frac{P_{Y^n}}{Q_{X^n}}}(y^n)\sqrt{\frac{P_{Y^n}(y^n)}{Q_{X^n}(y^n)}}\,Q_{X^n}(y^n)$$

$$= 1 - \sum_{y^n}\sqrt{P_{Y^n}(y^n)Q_{X^n}(y^n)}\sum_{x^n} T_i(x^n|y^n)\sqrt{\frac{P_{Y^n}(x^n)}{Q_{X^n}(x^n)}}$$

$$= 1 - \sum_{y^n}\sqrt{P_{Y^n}(y^n)Q_{X^n}(y^n)}\sum_{x_i} Q_{X_i|X_{i^c}}(x_i|y_{i^c})\sqrt{\frac{P_{Y_{i^c}}(y_{i^c})P_{Y_i|Y_{i^c}}(x_i|y_{i^c})}{Q_{X_{i^c}}(y_{i^c})Q_{X_i|X_{i^c}}(x_i|y_{i^c})}}$$

$$= 1 - \sum_{y^n}\sqrt{P_{Y_i|Y_{i^c}}(y_i|y_{i^c})Q_{X_i|X_{i^c}}(y_i|y_{i^c})}\;P_{Y_{i^c}}(y_{i^c})\sum_{x_i}\sqrt{P_{Y_i|Y_{i^c}}(x_i|y_{i^c})Q_{X_i|X_{i^c}}(x_i|y_{i^c})}$$

$$= 1 - \sum_{y_{i^c}} P_{Y_{i^c}}(y_{i^c})\left(\sum_{y_i}\sqrt{P_{Y_i|Y_{i^c}}(y_i|y_{i^c})Q_{X_i|X_{i^c}}(y_i|y_{i^c})}\right)^2$$

$$= 1 - \underset{\text{wrt } P_{Y_{i^c}}}{\mathbb{E}}\left[\Bigg(\underbrace{\sum_{y_i\in\mathcal{X}}\sqrt{P_{Y_i|Y_{i^c}}(y_i|Y_{i^c})Q_{X_i|X_{i^c}}(y_i|Y_{i^c})}}_{BC(P_{Y_i|Y_{i^c}}(\cdot|Y_{i^c}),\,Q_{X_i|X_{i^c}}(\cdot|Y_{i^c}))}\Bigg)^2\right].$$

$\leftarrow$ Bhattacharyya coefficient (or Hellinger affinity)

$$\Rightarrow \mathcal{E}_{T}\left(\sqrt{\frac{P_{Y^n}}{Q_{X^n}}},\sqrt{\frac{P_{Y^n}}{Q_{X^n}}}\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Bigg[\underbrace{1-\Bigg(\sum_{y\in\mathcal{X}}\sqrt{P_{Y_i|Y_{i^c}}(y|Y_{i^c})Q_{X_i|X_{i^c}}(y|Y_{i^c})}\Bigg)^2}_{\geq}\Bigg]$$

$$\geq \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left\|P_{Y_i|Y_{i^c}}(\cdot|Y_{i^c}) - Q_{X_i|X_{i^c}}(\cdot|Y_{i^c})\right\|_{TV}^2\right] \longleftarrow \left[\underline{\text{Lemma:}}\ \|P-Q\|_{TV}^2 \leq 1 - BC(P,Q)^2\right]$$

$$\geq \frac{\alpha}{4Cn} D(P_{Y^n}\|Q_{X^n}).$$ [using KL→TV tensorization result in Theorem 1]

∎

Def: (Dobrushin's Uniqueness Condition – $\mathcal{L}^2$ version)

$Q_{X^n} \in \mathcal{P}(\mathcal{X}^n)$ satisfies an $\mathcal{L}^2$-version of Dobrushin's uniqueness condition with coupling matrix $A = \{a_{\underset{\text{row}}{k},\underset{\text{column}}{i}}\}_{k,i=1}^{n}$, if for any $k,i \in [n]$ with $k \neq i$, and any $z^n, s^n \in \mathcal{X}^n$ with $z_{k^c} = s_{k^c}$ & $z_k \neq s_k$, we have:

differ only in kth coordinate

$$\left\|Q_{X_i|X_{i^c}}(\cdot|z_{i^c}) - Q_{X_i|X_{i^c}}(\cdot|s_{i^c})\right\|_{TV} \leq a_{k,i},$$

and setting $a_{ii} = 0$ for all $i \in [n]$, we have $\|A\|_2 \triangleq \max_{x:\|x\|_2=1}\|Ax\|_2 < 1$.

operator norm = largest singular value

Remark: In Dobrushin's original condition, we assume $\|A\|_1 \triangleq \max_{x:\|x\|_1=1}\|Ax\|_1 < 1$.

maximum absolute column sum

[continued.]

Anuran Makur
3 May 2017

Theorem 2: (Dobrushin's condition $\Rightarrow W_2^2$ tensorization)

If $Q_{X^n} \in \mathcal{P}(\mathcal{X}^n)$ satisfies Dobrushin's uniqueness condition with coupling matrix A with $\|A\|_2 < 1$, then for any $P_{Y^n} \in \mathcal{P}(\mathcal{X}^n)$:

cond.s of Thm 1
$$W_2^2(P_{Y_I|Y_{I^c}}(\cdot|y_{I^c}), Q_{X_I|X_{I^c}}(\cdot|y_{I^c})) \leq C \sum_{i\in I} \mathbb{E}\left[\left\|P_{Y_i|Y_{i^c}}(\cdot|Y_{i^c}) - Q_{X_i|X_{i^c}}(\cdot|Y_{i^c})\right\|_{TV}^2 \,\middle|\, Y_{I^c} = y_{I^c}\right]$$

for all $I \subseteq [n]$ and all $y_{I^c} \in \mathcal{X}^{[n]\setminus I}$, where $C = \dfrac{1}{(1-\|A\|_2)^2}$.

Proof: It suffices to prove this for $I = [n]$. (For any $I \subseteq [n]$ and any $y_{I^c} \in \mathcal{X}^{[n]\setminus I}$, $Q_{X_I|X_{I^c}}(\cdot|y_{I^c})$ satisfies Dobrushin's uniqueness condition with a minor of A, $\{a_{k,i}\}_{k,i\in I} = A_I$, as coupling matrix. Moreover, $\underset{\uparrow}{A_I^e} = \underset{\uparrow}{D_1} A \underset{\uparrow}{D_2} \Rightarrow \|A_I\|_2 = \|A_I^e\|_2 = \|D_1 A D_2\|_2 \leq \underbrace{\|D_1\|_2}_{\leq 1} \|A\|_2 \underbrace{\|D_2\|_2}_{\leq 1} \leq \|A\|_2$,

extend A_I with zero rows and columns so it is $n\times n$

diagonal matrices with 1's at I and 0's elsewhere

which implies that $\dfrac{1}{(1-\|A\|_2)^2} \geq \dfrac{1}{(1-\|A_I\|_2)^2}$. So, $C = 1/(1-\|A\|_2)^2$ is a valid constant for $I \subseteq [n]$.)

Idea: Prove that Gibbs sampler T (depends on Q_{X^n}) is a contraction wrt W_2-distance, and then use relationship between W_2 and TV distances.

Fix two distributions $R_{U^n}, S_{Z^n} \in \mathcal{P}(\mathcal{X}^n)$. Let π be a coupling between them.
Construct random variables $(U_1',\dots,U_n'), (Z_1',\dots,Z_n')$ as follows:

1. Sample a random variable $t \sim \text{Unif}[n]$ to get $t = i$, (t indep. of everything else)
2. Let $U_k' = U_k$ and $Z_k' = Z_k$ for $k \neq i$ so that the distribution of U_{i^c}', Z_{i^c}' given $t = i$ is $\pi_{U_{i^c}, Z_{i^c}}$,
3. Let the distribution of U_i', Z_i' given $t = i$ and $U_{i^c}' = u_{i^c}$, $Z_{i^c}' = z_{i^c}$ be the maximal coupling of $\underset{\uparrow}{Q_{X_i|X_{i^c}}(\cdot|u_{i^c})}$ and $\underset{\uparrow}{Q_{X_i|X_{i^c}}(\cdot|z_{i^c})}$ that achieves $\|Q_{X_i|X_{i^c}}(\cdot|u_{i^c}) - Q_{X_i|X_{i^c}}(\cdot|z_{i^c})\|_{TV}$.

   dist. of U_i' given $U_{i^c}' = u_{i^c}, t = i$; dist. of Z_i' given $Z_{i^c}' = z_{i^c}, t = i$

Clearly, the distribution of U'^n is $R_{U^n}T$ and the distribution of Z'^n is $S_{Z^n}T$.

For any $j \in [n]$,

$$\begin{aligned}
\mathbb{P}(U_j' \neq Z_j') &= \mathbb{P}(t\neq j)\,\mathbb{P}(U_j' \neq Z_j' | t \neq j) + \mathbb{P}(t=j)\,\mathbb{P}(U_j' \neq Z_j' | t = j) \\
&= \left(1-\tfrac{1}{n}\right) \underset{\text{wrt } \pi}{\mathbb{P}(U_j \neq Z_j)} + \frac{1}{n}\sum_{u_{j^c}, z_{j^c}} \underbrace{\mathbb{P}(U_j' \neq Z_j' | t=j, U_{j^c}' = u_{j^c}, Z_{j^c}' = z_{j^c})}_{= \|Q_{X_j|X_{j^c}}(\cdot|u_{j^c}) - Q_{X_j|X_{j^c}}(\cdot|z_{j^c})\|_{TV}} \underbrace{\mathbb{P}(U_{j^c}' = u_{j^c}, Z_{j^c}' = z_{j^c} | t = j)}_{= \pi_{U_{j^c}, Z_{j^c}}(u_{j^c}, z_{j^c})} \\
&= \left(1-\tfrac{1}{n}\right)\mathbb{P}(U_j \neq Z_j) + \frac{1}{n}\mathbb{E}\left[\|Q_{X_j|X_{j^c}}(\cdot|U_{j^c}) - Q_{X_j|X_{j^c}}(\cdot|Z_{j^c})\|_{TV}\right] \\
&\leq \left(1-\tfrac{1}{n}\right)\mathbb{P}(U_j \neq Z_j) + \frac{1}{n}\mathbb{E}\left[\sum_{k\neq j} \mathbb{1}\{U_k \neq Z_k\} a_{k,j}\right] \quad \leftarrow \text{use triangle inequality and Dobrushin's condition} \\
&= \left(1-\tfrac{1}{n}\right)\mathbb{P}(U_j \neq Z_j) + \frac{1}{n}\sum_{k\neq j} a_{k,j}\,\mathbb{P}(U_k \neq Z_k).
\end{aligned}$$

[continued.]

Anuran Makur
3 May 2017

Proof continued:

Let $p = [\mathbb{P}(U_1 \neq Z_1) \cdots \mathbb{P}(U_n \neq Z_n)]$ and $p' = [\mathbb{P}(U_1' \neq Z_1') \cdots \mathbb{P}(U_n' \neq Z_n')]$.

Then, we have:

$$\underbrace{p' \leq}_{\text{entrywise}} p\Big[\big(1-\tfrac{1}{n}\big)\underbrace{I}_{\text{identity matrix}} + \tfrac{1}{n}\underbrace{A}_{\text{diagonal entries}=0}\Big] \Rightarrow \|p'\|_2 \underbrace{\leq}_{\text{definition of } \|\cdot\|_2 \text{ for matrix}} \|p\|_2 \Big\|\big(1-\tfrac{1}{n}\big)I + \tfrac{1}{n}A\Big\|_2 \underbrace{\leq}_{\Delta\text{-inequality}} \Big(1 - \tfrac{1}{n}(1 - \|A\|_2)\Big)\|p\|_2$$

$$\Rightarrow \sqrt{\sum_{j=1}^{n} \mathbb{P}(U_j' \neq Z_j')^2} \leq \Big(1 - \tfrac{1}{n}(1-\|A\|_2)\Big) \sqrt{\sum_{j=1}^{n} \underbrace{\mathbb{P}}_{\text{wrt } \pi}(U_j \neq Z_j)^2}.$$

Choosing the minimum W_2-couplings, we get:

$$W_2(R_{U^n}T, S_{Z^n}T) \leq \Big(1 - \tfrac{1}{n}(1-\|A\|_2)\Big) W_2(R_{U^n}, S_{Z^n}). \quad [\text{contraction of } T \text{ in } W_2\text{-distance}]$$

By triangle inequality, for any $P_{Y^n} \in \mathcal{P}(\mathcal{X}^n)$:

$$W_2(P_{Y^n}, Q_{X^n}) \leq W_2(P_{Y^n}, P_{Y^n}T) + W_2(P_{Y^n}T, \overbrace{Q_{X^n}}^{= Q_{X^n}T \text{ [invariant measure]}})$$
$$\leq W_2(P_{Y^n}, P_{Y^n}T) + \Big(1 - \tfrac{1}{n}(1-\|A\|_2)\Big) W_2(P_{Y^n}, Q_{X^n}) \quad [\text{contraction}]$$

$$\Rightarrow W_2(P_{Y^n}, Q_{X^n}) \leq \frac{n}{1-\|A\|_2} W_2(P_{Y^n}, P_{Y^n}T). \quad [*]$$

Now we upper bound $W_2(P_{Y^n}, P_{Y^n}T)$. We construct a coupling of $P_{Y^n}, P_{Y^n}T \in \mathcal{P}(\mathcal{X}^n)$ using random variables Y^n, Y'^n:

1. Sample a random variable $t \sim \text{Unif}[n]$ to get $t = i$, (t indep. of everything else)
2. Let $Y_k = Y_k'$ for $k \neq i$ so that the distribution of Y_{i^c} given $t=i$ is $P_{Y_{i^c}}$,
3. Let the distribution of Y_i, Y_i' given $t=i$ and $Y_{i^c} = y_{i^c}$ be the maximal coupling of $\underbrace{P_{Y_i|Y_{i^c}}(\cdot|y_{i^c})}_{\text{dist. of } Y_i \text{ given } t=i \text{ and } Y_{i^c}=y_{i^c}}$ and $\underbrace{Q_{X_i|X_{i^c}}(\cdot|y_{i^c})}_{\text{dist. of } Y_i' \text{ given } t=i \text{ and } Y_{i^c}=y_{i^c}}$ that achieves $\|P_{Y_i|Y_{i^c}}(\cdot|y_{i^c}) - Q_{X_i|X_{i^c}}(\cdot|y_{i^c})\|_{TV}$.

Clearly, the distribution of Y^n is P_{Y^n} and the distribution of Y_n' is $P_{Y^n}T$.

For any $j \in [n]$,

$$\mathbb{P}(Y_j \neq Y_j') = \frac{1}{n}\mathbb{P}(Y_j \neq Y_j' | t=j) + \Big(1-\frac{1}{n}\Big)\overbrace{\mathbb{P}(Y_j \neq Y_j' | t \neq j)}^{=0}$$
$$= \frac{1}{n}\sum_{y_{j^c}} \mathbb{P}(Y_j \neq Y_j' | t=j, Y_{j^c} = y_{j^c})\, \mathbb{P}(Y_{j^c} = y_{j^c} | t=j)$$
$$= \frac{1}{n}\mathbb{E}_{P_{Y_{j^c}}}\Big[\|P_{Y_j|Y_{j^c}}(\cdot|Y_{j^c}) - Q_{X_j|X_{j^c}}(\cdot|Y_{j^c})\|_{TV}\Big]$$

$$\Rightarrow \mathbb{P}(Y_j \neq Y_j')^2 = \frac{1}{n^2}\mathbb{E}\Big[\|P_{Y_j|Y_{j^c}}(\cdot|Y_{j^c}) - Q_{X_j|X_{j^c}}(\cdot|Y_{j^c})\|_{TV}\Big]^2 \underbrace{\leq}_{\text{Jensen's inequality}} \frac{1}{n^2}\mathbb{E}\Big[\|P_{Y_j|Y_{j^c}}(\cdot|Y_{j^c}) - Q_{X_j|X_{j^c}}(\cdot|Y_{j^c})\|_{TV}^2\Big].$$

Hence, we have: $W_2(P_{Y^n}, P_{Y^n}T) \leq \sqrt{\sum_{i=1}^{n}\mathbb{P}(Y_i \neq Y_i')^2} \leq \frac{1}{n}\sqrt{\sum_{i=1}^{n}\mathbb{E}\Big[\|P_{Y_i|Y_{i^c}}(\cdot|Y_{i^c}) - Q_{X_i|X_{i^c}}(\cdot|Y_{i^c})\|_{TV}^2\Big]}$,

which with [*] completes the proof. ∎

~◎~

[THE END]