These notes have not received the scrutiny of publication. They could be missing important references, etc.

Dobrushin's Condition

1 Review

1.1 What We're at

Let π be a probability measure and we want to sample from it. P is the transition matrix s.t. $\pi P = \pi$ and the process is reversible and ergodic. Now we know that:

- $d_{\chi^2}(\tau P^t, \pi) \le (\max_i |\lambda_i|)^t \, d_{\chi^2}(\tau, \pi)$

  We can always ensure that $\max_i |\lambda_i| = \lambda_2$ by letting $P' = \frac{1}{2}(I + P)$.

- $\lambda_2(P)$ is the smallest λ_2 s.t. $\mathrm{Var}_\pi(f) \le \frac{1}{1-\lambda_2}\langle f, (I-P)f\rangle_\pi$.

Here're some examples:

- SRW on s-cycle: $1 - \lambda_2 = \mathcal{O}\left(\frac{1}{s^2}\right)$
- Gibbs sampler on a prob measure $\pi = \pi_1 \otimes \cdots \otimes \pi_n$ measure on $\mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n$: e.g. $\{\pm 1\}^n$

  By Efron-Stein's formula, we know that $\lambda_2 = \frac{n-1}{n} = 1 - \frac{1}{n} = 1 - \frac{1}{\log_2 s}$, where $s = \#$ states $= 2^n$. Then $1 - \lambda_2 = \frac{1}{\log_2 s}$.

1.2 What's Next

- Gibbs Sampler under weak dependence: "Dobrushin's condition"

2 Dobrushin Influence Matrix

Let $\pi(x_1, ..., x_n)$ be a prob measure on $\bigotimes_{i=1}^n \mathcal{E}_i$, where $x_i \in \mathcal{E}_i$. Define the Dobrushin Influence Matrix be $R \in \mathbb{R}^{n\times n}$, where

$$R_{ij} = \max_{\substack{y,z \\ y_{\sim j} = z_{\sim j}}} |\pi(x_i = \cdot | x_{\sim i} = y_{\sim i}) - \pi(x_i = \cdot | x_{\sim i} = z_{\sim i})|_{\mathrm{TV}}.$$

Remark 1. $|\cdot|_{\mathrm{TV}}$ stands for the Total Variation distance where $|p-q|_{\mathrm{TV}} = d_{\mathrm{TV}}(p,q) = \frac{1}{2}\sum_x |p(x) - q(x)|$ for all probability measure p, q.

Remark 2. $\forall i, R_{ii} = 0$.

Remark 3. If π is a product measure, that is, all elements are independent, then $R = \mathbf{0}^{n\times n}$. That's because $R_{ij} = \max_{\substack{y,z \\ y_{\sim j} = z_{\sim j}}} |\pi(x_i = \cdot | x_{\sim i} = y_{\sim i}) - \pi(x_i = \cdot | x_{\sim i} = z_{\sim i})|_{\mathrm{TV}} = \max_{\substack{y,z \\ y_{\sim j} = z_{\sim j}}} |\pi(x_i = \cdot) - \pi(x_i = \cdot)|_{\mathrm{TV}} = 0$ for all i, j.

3 Spectral Gap Theorem [Hayes '05 and Wu '06]

Theorem 1. *Let P be the Gibbs sampler for π, and R is the corresponding Dobrushin Influence Matrix, then $1-\lambda_2(P) \geq \frac{1-|R|_{op}}{n}$, where $|R|_{op}$ =largest singular value of $R = \max_{|w|_2\leq 1} |Rw|_2$.*

Corollary 1 (Ising model)**.** *The probability measure $\pi(x) = \frac{1}{z}\exp(\frac{1}{2}\langle x, Jx\rangle)$, $x \in \{\pm 1\}^n$. As $\mathbb{E}[x_i|x_{\sim i}] = \tanh(J_{\sim i}x_{\sim i})$, we have*

$$\begin{aligned} R_{ij} &= \frac{1}{2}\max_{y_{\sim j}=z_{\sim j}} |\tanh(J_{i,\sim i}y_{\sim i}) - \tanh(J_{i,\sim i}z_{\sim i})| \\ &\leq \frac{1}{2}\max_{y_{\sim j}=z_{\sim j}} |\langle J_{i,\sim i}\ ,\ y_{\sim i} - z_{\sim i}\rangle| \\ &= \frac{1}{2}\max_{y_j=z_j} |J_{ij}||y_j - z_j| \\ &= |J_{ij}|. \end{aligned}$$

Then

$$1-\lambda_2(P) \geqslant \frac{1-\left|(|J_{ij}|)_{ij}\right|_{op}}{n}.$$

If we require $J_{ij} \geqslant 0$, we will have

$$1-\lambda_2(P) \geqslant \frac{1-|J|_{op}}{n}.$$

Proof. (of Corollary 1) We know that $R_{ij} \geq 0$ always. For the case $J_{ij} \geq 0$, both R and J have a unique maximal eigenvalue with eigenvector that has all positive entries according to Perron-Frobenius theorem. Therefore,

$$|R|_{op} = \max_{\substack{|u|_2=1 \\ u_i\geq 0,\ \forall i}} |Ru|_2 \leq \max_{\substack{|u|_2=1 \\ u_i\geq 0,\ \forall i}} |Ju|_2 = |J|_{op}.$$

□

Example 1. In Ising model, let $J = \beta A$, where $\beta \geq 0$ is the "inverse temperature" and A is the adjacency matrix of G. Let d = max degree of graph $G = \max_i \sum_j A_{ij}$. Then $|J|_{op} = \beta|A|_{op} \leq \beta d$. By Corollary 1, we have

$$1-\lambda_2(P) \geq 1 - \frac{\beta d}{n}.$$

Remark 4. Previously if $P \neq NP$ then there's no polytime algorithm to compute $z = \sum_x e^{\frac{\beta\langle x, Ax\rangle}{2}}$ even when $\beta d < \frac{1}{100}$. This is called "exact counting" and it's generally impossible. However, $\forall\epsilon > 0$, computing $(1\pm\epsilon)z$ in poly $\left(n, \frac{1}{c}\right)$ time is feasible for $\beta d < 0.99$. This is called "approximate counting".

Definition 1 (Marton's Wasserstein distance)**.** The Marton's Wasserstein distance is defined as $W^2(\lambda,\tau) = \inf_{\rho(y,z)} \sum_{i=1}^n \rho^2(y_i \neq z_i)$, where the infimum ranges over all couplings ρ of λ and τ, i.e. joint distributions $\rho(y,z)$ with marginals $\rho|_y = \lambda$ and $\rho|_z = \tau$. And note $W(\lambda,\tau) = \sqrt{W^2(\lambda,\tau)}$.

Remark 5. For y, z that have same marginal distribution λ, we have $W^2(\lambda,\lambda) = 0$.

Lemma 1. *Under setting of Theorem 1, we have $W(\lambda P, \tau P) \leq \left(1 - \frac{1-|R|_{op}}{n}\right) W(\lambda.\tau)$, for all probability measures λ, τ.*

Proof. (of Lemma 1)

Step 1. Construct coupling:

(1) Let ρ be the minimizer of $W(\lambda,\tau)$;

(2) Sample (y,z) $\sim \rho$;

(3) Pick $i \sim \mathrm{Unif}([n])$ (due to procedures in Gibbs sampler P);

(4) Sample $u \sim \mathrm{Unif}[0,1]$;

(5) Let $y_i' = \begin{cases} 1 & \text{if } u < \pi\left(x_i = 1 \mid x_{\sim i} = y_{\sim i}\right) \\ -1 & \text{o.w.} \end{cases}$, $z_i' = \begin{cases} 1 & \text{if } u < \pi\left(x = 1 \mid x_{\sim i} = z_{\sim i}\right) \\ -1 & \text{o.w.} \end{cases}$,
$y'_{\sim i} = y_{\sim i}$, and $z'_{\sim i} = z_{\sim i}$;

(6) Let ρ' be the joint law of (y', z').

Step 2. Prove that $\rho'\left(y_i' \neq z_1'\right) \leqslant \frac{n-1}{n}\rho\left(y_i \neq z_i\right) + \frac{1}{n}\sum_{j\neq i} R_{ij}\rho\left(y_j \neq z_j\right)$, where ρ' is constructed from Step 1.

$$\begin{aligned}\rho'\left(y_i' \neq z_i'\right) = &\frac{n-1}{n}\rho\left(y_i \neq z_i\right) \quad \text{[don't update } i\text{]} \\ &+ \frac{1}{n}\mathbb{E}_\rho|\pi(x_i = \cdot \mid x_{\sim i} = y_{\sim i}) - \pi(x_i = \cdot \mid x_{\sim i} = z_{\sim i})|_{\mathrm{TV}} \quad \text{[do update } i\text{]}\end{aligned}$$

Notice that the second term satisfies

$$\begin{aligned}&|\pi(x_i = \cdot \mid x_{\sim i} = y_{\sim i}) - \pi(x_i = \cdot \mid x_{\sim i} = z_{\sim i})|_{\mathrm{TV}} \\ \leqslant &\sum_{j\neq i}^{n} |\pi\left(x_i = \cdot \mid x_{<j} = y_{<j}, x_{>j} = z_{>j},\ \text{skip } i\right) - \pi(x_i = \cdot \mid x_{\leqslant j} = y_{\leqslant j}, x_{>j} = z_{>j},\ \text{skip } i)| \\ \leqslant &\sum_{j\neq i}^{n} R_{ij}\mathbb{1}\left(y_j \neq z_j\right).\end{aligned}$$

By pluging in back, we derive that

$$\begin{aligned}\rho'\left(y_i' \neq z_1'\right) &\leq \frac{n-1}{n}\rho\left(y_i \neq z_i\right) + \frac{1}{n}\sum_{j\neq i}^{n} R_{ij}\mathbb{E}_\rho\mathbb{1}\left(y_j \neq z_j\right) \\ &\leq \frac{n-1}{n}\rho\left(y_i \neq z_i\right) + \frac{1}{n}\sum_{j\neq i} R_{ij}\rho\left(y_j \neq z_j\right).\end{aligned}$$

Step 3. Let $r_i' = \rho'\left(y_i' \neq z_i'\right), r_i = \rho\left(y_i \neq z_i\right)$, then $r_i' = \frac{n-1}{n}r_i + \frac{1}{n}(Rr)_i$ and as a vector

$$\begin{aligned}|r'|_2 &= \left|\frac{n-1}{n}r + \frac{1}{n}Rr\right|_2 \\ &= \left|\left(\frac{n-1}{n}I + \frac{1}{n}R\right)r\right|_2 \\ &\leqslant \left|\frac{n-1}{n}I + \frac{1}{n}R\right|_{\mathrm{op}} |r|_2 \\ &\leqslant \left(\frac{n-1}{n} + \frac{|R|_{\mathrm{op}}}{n}\right)|r|_2.\end{aligned}$$

That is, $W(\lambda P, \tau P) \leqslant \left(\frac{n-1}{n} + \frac{|R|_{op}}{n}\right) W(\lambda, \tau)$.

□

Proof. (of Theorem 1) Lemma 1 implies some claims for d_{χ^2}, let $f_2 = 2^{\text{nd}}$ largest eigenvector of P, then we want to show

$$\begin{aligned} \left|(\lambda-\tau)P^t f_2\right| &= \left(\max_{i\neq 1}|\lambda_i|\right)^t |(\lambda-\tau)f_2| \\ &\overset{boring}{\leq} 2\sqrt{n}W(\lambda,\tau)|f|_\infty \left(1-\frac{1-|R|_{op}}{n}\right)^t. \end{aligned}$$

□