Constraint Relaxation for Bayesian Modeling with Parameter Constraints

Leo Duan, Alex Young, Akihiko Nishimura, David Dunson

Assume the approximating function for $\{v_j(\theta)=0\}_{j=1}^s$ take the form $\exp(-\frac{\|v(\theta)\|_1}{\lambda}) = \exp(-\frac{\sum_j |v_j(\theta)|}{\lambda})$, with $\lambda>0$ a scalar. And $\mathcal{R}$ and $\{v_j(\theta)\}_{j=1}^s$ satisfy $\int \mathbb{1}_{v(\theta)=x}\pi_{\mathcal{R}}(\theta)\bar{\mathcal{H}}^{p-s}(d\theta) \in (0,\infty)$ for any $x\in\mathcal{X} = \{v(\theta);\theta\in\mathcal{R}\}$. We denote conditional expectation $\mathbb{E}(g(\theta)\mid v(\theta)=x) = \mathbb{E}(g(\theta)\mid x) = \int_{v^{-1}(x)} g(\theta)\frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)}\bar{\mathcal{H}}^{p-s}(d\theta)$.

Remark 1. *The 1-Wasserstein distance between the measures based on* (**??**) *and* (**??**) *has*

$$\lim_{\lambda\to 0} W_1(\Pi,\tilde{\Pi}) = 0.$$

Further, for $\alpha=1$ in (**??**),

$$W_1(\Pi,\tilde{\Pi}) \le (2\lambda)^s\left(\frac{k_1k_3}{m_0^2}+\frac{k_1}{m_0}\right) + \exp(-\lambda^{-1}t)\left(\frac{k_1}{m_0^2}+\frac{k_2}{m_0}\right), \tag{1}$$

where $k_1 = \sup_{g:\|g\|_L\le 1}\ \sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}} \|\mathbb{E}(g(\theta)\mid x^*)\|$, $k_2 = \sup_{g:\|g\|_L\le 1}\mathbb{E}(\|g(\theta)\|\mathbb{1}_{v(\theta)\in\mathcal{X}})$, $k_3 = \sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}} m(x^*)$.

Proof. The co-area formula from Federer (2014) is,

$$\int_{\mathbb{R}^p} f(\theta)J\Phi(\theta)\mu^p(d\theta) = \int_{\mathbb{R}^s}\left[\int_{\Phi^{-1}(x)} f(\theta)\bar{\mathcal{H}}^{p-s}(d\theta)\right]\mu^s(dx), \tag{2}$$

where $\mu^k(d\theta)$ a k-dimensional Lebesgue measure and $\bar{\mathcal{H}}^s(d\theta)$ is a s-dimensional normalized Hausdorff measure.

Let $g:\mathbb{R}^p\to\mathbb{R}$ be a 1-Lipschitz continuous function, i.e. $\|g(x)-g(y)\|\le\|x-y\|$, denoted by $\|g\|_L\le 1$. By Kantorovich-Rubinstein duality, the 1-Wasserstein distance based on Euclidean metric equals to:

$$W_1(\Pi,\tilde{\Pi}) = \sup_{g:\|g\|_L\le 1}\int g(x)\Pi(dx) - \int g(y)\tilde{\Pi}(dy) \tag{3}$$

By assumption, $\pi_{\mathcal{R}}(\theta) = \pi_{\mathcal{R}}(\theta)\mathbb{1}_{v(\theta)\in\mathcal{X}}$. Taking $f(\theta) = \frac{\exp(-\lambda^{-1}\|v(\theta)\|_1)\pi_{\mathcal{R}}(\theta)\mathbb{1}_{v(\theta)\in\mathcal{X}}}{Jv(\theta)}$ and $\Phi(\theta)=v(\theta)$ in the co-area formula yields

$$
\begin{aligned}
m_\lambda &= \int_{\mathbb{R}^s} \left[\int_{v^{-1}(x)} \frac{\exp(-\lambda^{-1}\|v(\theta)\|_1)\pi_{\mathcal{R}}(\theta)}{Jv(\theta)} \bar{\mathcal{H}}^{p-s}(d\theta) \right] \mathbb{1}_{x\in\mathcal{X}} \mu^s(dx) \\
&= \int_{\mathbb{R}^s} \left[\int_{v^{-1}(x)} \frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)} \bar{\mathcal{H}}^{p-s}(d\theta) \right] \exp(-\lambda^{-1}\|x\|_1) \mathbb{1}_{x\in\mathcal{X}} \mu^s(dx) \qquad (4) \\
&= \int_{\mathbb{R}^s} m(x) \exp(-\lambda^{-1}\|x\|_1) \mathbb{1}_{x\in\mathcal{X}} \mu^s(dx).
\end{aligned}
$$

Taking $f(\theta) = \frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)} \mathbb{1}_{v(\theta)=\mathbf{0}}$ and $\Phi(\theta) = v(\theta)$ yields

$$
m_0 = \int_{\mathbb{R}^s} \left[\int_{v^{-1}(x)} \frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)} \bar{\mathcal{H}}^{p-s}(d\theta) \right] \mathbb{1}_{x=\mathbf{0}} \mu^s(dx) = \int_{v^{-1}(\mathbf{0})} \frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)} \bar{\mathcal{H}}^{s}(d\theta) = m(\mathbf{0}) \qquad (5)
$$

Clearly $m_\lambda \geq m_0$.

1. Asymptotic result:

$$
\begin{aligned}
&\sup_{g:\|g\|_L\leq 1} \int_{\mathbb{R}^s} \int_{v^{-1}(x)} g(\theta) \left[\frac{\exp(-\lambda^{-1}\|v(\theta)\|_1)}{m_\lambda} - \frac{\mathbb{1}_{v(\theta)=\mathbf{0}}}{m_0} \right] \frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)} \bar{\mathcal{H}}^{p-s}(d\theta) \mathbb{1}_{x\in\mathcal{X}} \mu^s(dx) \\
&= \sup_{g:\|g\|_L\leq 1} \int_{\mathbb{R}^s} \mathbb{E}(g(\theta) \mid x) \left[\frac{\exp(-\lambda^{-1}\|x\|_1)}{m_\lambda} - \frac{\mathbb{1}_{x=\mathbf{0}}}{m_0} \right] \mathbb{1}_{x\in\mathcal{X}} \mu^s(dx) \\
&= \sup_{g:\|g\|_L\leq 1} \int_{\mathbb{R}^s} \mathbb{E}(g(\theta) \mid x) \left[\frac{1}{m_\lambda} - \frac{1}{m_0} \right] \mathbb{1}_{x=\mathbf{0}} \mu^s(dx) + \sup_{g:\|g\|_L\leq 1} \int_{\mathbb{R}^s} \mathbb{E}(g(\theta) \mid x) \frac{\exp(-\lambda^{-1}\|x\|_1)}{m_\lambda} \mathbb{1}_{x\neq\mathbf{0}} \mathbb{1}_{x\in\mathcal{X}} \mu^s(dx) \\
&\leq \sup_{g:\|g\|_L\leq 1} \|\mathbb{E}(g(\theta) \mid \mathbf{0})\| \left[\frac{1}{m_0} - \frac{1}{m_\lambda} \right] + \frac{1}{m_0} \sup_{g:\|g\|_L\leq 1} \int_{\mathbb{R}^s} \|\mathbb{E}(g(\theta) \mid x)\| \exp(-\lambda^{-1}\|x\|_1) \mathbb{1}_{x\neq\mathbf{0}} \mathbb{1}_{x\in\mathcal{X}} \mu^s(dx)
\end{aligned}
\qquad (6)
$$

Note $m_\lambda \leq \int_{\mathbb{R}^s} m(x) \mathbb{1}_{x\in\mathcal{X}} \mu^s(dx) = \int_{\mathbb{R}^p} \pi_{\mathcal{R}}(\theta) d\theta = 1$. By dominated convergence theorem,

$$
\lim_{\lambda\to 0} m_\lambda = \int_{\mathbb{R}^s} m(x) \lim_{\lambda\to 0} \exp(-\lambda^{-1}\|x\|_1) \mathbb{1}_{x\in\mathcal{X}} \mu^s(dx) = m_0. \qquad (7)
$$

Since

$$
\sup_{g:\|g\|_L\leq 1} \int_{\mathbb{R}^s} \|\mathbb{E}(g(\theta) \mid x)\| \exp(-\lambda^{-1}\|x\|_1) \mathbb{1}_{x\neq\mathbf{0}} \mathbb{1}_{x\in\mathcal{X}} \mu^s(dx) \leq \int_{\mathbb{R}^s} \sup_{g:\|g\|_L\leq 1} \|\mathbb{E}(g(\theta) \mid x)\| \exp(-\lambda^{-1}\|x\|_1) \mathbb{1}_{x\neq\mathbf{0}} \mathbb{1}_{x\in\mathcal{X}} \mu^s(dx),
$$

letting $q_\lambda = \sup_{g:\|g\|_L \leq 1} \|\mathbb{E}(g(\theta) \mid x)\| \exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\neq \mathbf{0}}\mathbb{1}_{x\in\mathcal{X}}$, for fixed x, we have $0 \leq q_1 - q_{\lambda_1} \leq q_1 - q_{\lambda_2}$ for any two numbers in the series $1 \geq \lambda_1 \geq \lambda_2$, by monotone convergence theorem, $\lim_{\lambda\to 0}\int[q_1(x) - q_\lambda(x)]dx = \int[q_1(x) - q_0(x)]dx$ hence $\lim_{\lambda\to 0}\int q_\lambda(x)dx = 0$. Combining the results yields

$$\lim_{\lambda\to 0} W_1(\Pi, \tilde{\Pi}) = 0. \tag{8}$$

2. Non-asymptotic result:

$$\begin{aligned}
\frac{1}{m_0} - \frac{1}{m_\lambda} &\leq \frac{\int_{\mathbb{R}^s} m(x)\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\neq\mathbf{0}}\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)}{m_0^2} \\
&= \frac{1}{m_0^2}\left[\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(0,t]}m(x)\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx) + \int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(t,\infty)}m(x)\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\right] \\
&\leq \frac{1}{m_0^2}\Bigg[\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}} m(x^*)\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(0,t]}\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx) \\
&\quad + \exp(-\lambda^{-1}t)\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(t,\infty]}m(x)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\Bigg] \\
&\leq \frac{1}{m_0^2}\left[(2\lambda)^s\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}} m(x^*) + \exp(-\lambda^{-1}t)\right]
\end{aligned} \tag{9}$$

where

$$\begin{aligned}
\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(0,t]}\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx) &= \int_{\mathbb{R}^s}\mathbb{1}_{x\in\mathcal{X}}\mathbb{1}_{\|x\|_1\in(0,t]}\prod_{i=1}^s\exp(-\lambda^{-1}|x_i|)\mu^s(dx) \\
&\leq \int_{\mathbb{R}^s}\prod_{i=1}^s\exp(-\lambda^{-1}|x_i|)\mu^s(dx) \\
&= (2\lambda)^s
\end{aligned} \tag{10}$$

$$
\begin{aligned}
&\sup_{g:\|g\|_L\le 1}\int_{\mathbb{R}^s}\|\mathbb{E}(g(\theta)\mid x)\|\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\neq\mathbf{0}}\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\\
&\le \sup_{g:\|g\|_L\le 1}\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}}\|\mathbb{E}(g(\theta)\mid x^*)\|\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(0,t]}\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\\
&+\exp(-\lambda^{-1}t)\sup_{g:\|g\|_L\le 1}\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(t,\infty)}\|\mathbb{E}(g(\theta)\mid x)\|\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\\
&\le \sup_{g:\|g\|_L\le 1}\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}}\|\mathbb{E}(g(\theta)\mid x^*)\|\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(0,t]}\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\\
&+\exp(-\lambda^{-1}t)\sup_{g:\|g\|_L\le 1}\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(t,\infty)}\mathbb{E}(\|g(\theta)\|\mid x)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\\
&\le \sup_{g:\|g\|_L\le 1}\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}}\|\mathbb{E}(g(\theta)\mid x^*)\|(2\lambda)^s+\exp(-\lambda^{-1}t)\sup_{g:\|g\|_L\le 1}\mathbb{E}(\|g(\theta)\|\mathbb{1}_{v(\theta)\in\mathcal{X}})
\end{aligned}
\tag{11}
$$

,

Combining (6)(9)(11), $k_1=\sup_{g:\|g\|_L\le 1}\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}}\|\mathbb{E}(g(\theta)\mid x^*)\|$, $k_2=\sup_{g:\|g\|_L\le 1}\mathbb{E}(\|g(\theta)\|\mathbb{1}_{v(\theta)\in\mathcal{X}})$, $k_3=\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}}m(x^*)$

$$
\begin{aligned}
&\sup_{g:\|g\|_L\le 1}\int g(x)\Pi(dx)-\int g(x)\tilde{\Pi}(dx)\\
&\le (2\lambda)^s(\frac{k_1k_3}{m_0^2}+\frac{k_1}{m_0})+\exp(-\lambda^{-1}t)(\frac{k_1}{m_0^2}+\frac{k_2}{m_0})
\end{aligned}
\tag{12}
$$

□

The first part shows the asymptotic accuracy of the approximation. The second part shows the rate with non-asymptotic λ under mild assumptions. The interpretation for these assumptions is that if in a small space expansion of $\mathcal{D}$, defined as $\{\theta^*:\|v(\theta^*)\|_1\in[0,t]\}$, the marginal density of $v(\theta^*)$ and the conditional expectation of Lipschitz functions are bounded $k_1,k_2=\mathcal{O}(1)$, and the expected norm of Lipschitz function are smaller than a bound that grows near exponentially $k_3=\mathcal{O}(\lambda\exp(t/\lambda))$, then the distance $W_1(\Pi,\tilde{\Pi})$ converges to 0 in $\mathcal{O}(\lambda^s)$ as $\lambda\to 0$.

References

Federer, H. (2014). *Geometric measure theory.* Springer.