Two possible smoothness assumptions

Miaoyan Wang, 03/29/2021.

1 Option 1. Uniform assumption over all finite d

We introduce a small tolerance $\Delta s = \Delta s(d)$ to account for discrete measure with finite d. Specifically, for each given tensor dimension $d \geq 2$, let $G_d \colon [-1,1] \to [0,1]$ be the empirical cumulative function of entries in Θ, and $\Delta s = \frac{1}{d^K}$ the tolerance point mass. We call π a mass point if its measure under G_d is heavier than than the tolerance mass. Denote

$$\mathcal{N}_d = [-1,1]/\left\{\pi \colon \inf_{0<t\leq \Delta s} \frac{G_d(\pi+t)-G_d(\pi-t)}{t} \leq c_1 < \infty\right\}$$

the set of mass points in G_d, where c_1 denotes a positive constant independent of tensor dimension d.

Assumption 1 (α-smoothness of all finite d)**.** Fix $\pi \notin \mathcal{N}_d$. Assume there exist positive constants $\alpha > 0, c > 0$, independent of tensor dimension d, such that

$$\sup_{\Delta s \leq t < \rho(\pi,\mathcal{N})} \frac{G_d(\pi+t)-G_d(\pi-t)}{t^\alpha} \leq c < \infty.$$

We make the conversion that $\alpha = \infty$ if $\rho(\pi, \mathcal{N}) \leq \Delta s$.

2 Option 2. Functional assumption for infinite d

Another approach is to construct a latent infinite-dimensional process under $(\Theta_d)_{d\geq 2}$. The idea is to use Lebesque representation of a measurable function from $\mathcal{X}$ to $\mathbb{R}$:

$$f \Longleftrightarrow \{\boldsymbol{x} \in \mathcal{X} \colon f(\boldsymbol{x}) \leq \pi\}, \quad \text{for all } \pi \in \mathbb{R}. \tag{1}$$

Because the correspondence is one-to-one except for a measure-zero set (?), we will refer to "implicit construction" of f using the right hand side of (1).

Definition 1 (α-smooth, r-sign representable function)**.** Let f be a multivariate function

$$\begin{aligned} f \colon [0,1]^K &\to [-1,1], \\ \boldsymbol{x} &\mapsto f(\boldsymbol{x}). \end{aligned}$$

We call f is a α-smooth, sign-r representable function if f satisfies the following two conditions:

1. (α-smoothness) Define a function $G(\pi) = \text{Leb}\,(\boldsymbol{x}\colon f(\boldsymbol{x}) \leq \pi)$, where $\text{Leb}(\cdot)$ denotes the Lebesgue measure in $\mathbb{R}^K$. Assume the function $G\colon [-1,1] \to [0,1]$ is α-smooth.

2. (r-sign representability) For each $d \geq 2$, let $\mathcal{D}_d\colon f \mapsto \Theta$ denote the degree-d interpolation of f that induces a dimensional-$(d,\ldots,d)$ tensor Θ, where

$$\Theta(i_1,\ldots,i_d) = f\left(\frac{i_1}{d},\ldots,\frac{i_K}{d}\right), \quad \text{for all } (i_1,\ldots,i_K) \in [d]^K.$$

   Assume $\sup_{d\geq 2}\sup_{\pi\in[-1,1]} \text{srank}(\mathcal{D}_d(f-\pi)) \leq r$ for some rank r.

We use $\mathcal{F}(r,\alpha)$ to denote the family of functions that satisfy the above two conditions.

Set-up:

1. Signal tensor $\Theta = \mathcal{D}_d(f)$, where $f \in \mathcal{F}(r,\alpha)$.

2. Observed tensor $\mathcal{Y} = \Theta + \mathcal{E}$. Assume complete observation for simplicity.

3. Estimated tensor $\hat{\Theta}$, and the associated empirical CDF $\hat{G}_d(\cdot)\colon [-1,1] \to [0,1]$,

$$\hat{G}_d(\pi) = \frac{\left|\left\{\omega \in [d]^K \colon \hat{\Theta}(\omega) \leq \pi\right\}\right|}{d^K}.$$

4. Estimated function $\hat{f}$. We implicitly define a function estimate $\hat{f}\colon [0,1]^K \to [-1,1]$ such that

$$\text{Leb}\{\boldsymbol{x} \in [0,1]^K \colon \hat{f}(\boldsymbol{x}) \leq \pi\} = \hat{G}_d(\pi), \quad \text{for all } \pi \in [-1,1]. \tag{2}$$

5. Estimation error. Under definition (2), we have

$$\|\hat{f} - f\|_1 := \underbrace{\int_{\boldsymbol{x}\in[0,1]^K} |\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})| d\boldsymbol{x}}_{\text{Riemann integral}} \stackrel{?}{=} \underbrace{\int_{\theta\in[-1,1]} |\hat{\theta} - \theta| dG(\theta)}_{\text{Lebesque integral}} =: \mathbb{E}_{\theta\sim G}|\hat{\theta} - \theta| = \text{MAE}(\hat{\theta},\theta), \tag{3}$$

where the evaluation of Lebesque integral is equivalent to viewing θ as a random variable from the distribution G in the infinite-dimensional process.

Remark 1 (Explicit construction of $\hat{f}$)**.** The integrated loss (3) relies on only the indicator function (2) of $\hat{f}$; the pointwise definition of $\hat{f}$ is unimportant. In principle, all our current proof should also go through because we only use the integral property of $\hat{f}$ (verify?).

Here we provide a constructive definition for completeness. We use $\boldsymbol{x} \in [0,1]^K$ to denote the continuous domain variable and $\omega \in [d]^K$ the discrete domain variable. Define the function $\hat{f}$ using a stepwise function

$$\hat{f}(\boldsymbol{x}) = \hat{\Theta}(\lfloor d\boldsymbol{x} \rfloor),$$

where $\lfloor\cdot\rfloor$ is applied to vectors in an entrywise fashion. Under this construction, the right hand side of (2) can be expressed as

$$\left\{\boldsymbol{x}\colon \hat{f}(\boldsymbol{x}) \le \pi\right\} = \left\{\boldsymbol{x}\colon \hat{\Theta}(\lfloor d\boldsymbol{x}\rfloor) \le \pi\right\}$$
$$= \bigcup_{\omega\in[d]^K} \left\{\omega \le d\boldsymbol{x} < \omega + (1,\dots,1)\colon \hat{\Theta}(\omega) \le \pi\right\}.$$

Therefore we have

$$\mathrm{Leb}\left(\boldsymbol{x}\colon \hat{f}(\boldsymbol{x}) \le \pi\right) = \sum_{\omega\in[d]^K} \mathbb{1}\{\omega\colon \hat{\Theta}(\omega) \le \pi\} \times \mathrm{Leb}\left(\boldsymbol{x} \in \frac{1}{d}[\omega, \omega + (1,\dots,1))\right)$$
$$= \sum_{\omega\in[d]^K} \frac{\mathbb{1}\{\omega\colon \hat{\Theta}(\omega) \le \pi\}}{d^K} = \hat{G}(\pi).$$