Optimal Rates for the Random Fourier Feature Technique

Zoltán Szabó

Joint work with Bharath K. Sriperumbudur (PSU)

École Polytechnique
March 14, 2016

Outline

- Kernels.
- Random Fourier features (RFFs).
- Guarantees on RFF approximation: uniform, L^p.

Kernel, RKHS

- $k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ kernel on $\mathcal{X}$, if
  - $\exists \varphi: \mathcal{X} \to H$(ilbert space) feature map,
  - $k(a,b) = \langle \varphi(a), \varphi(b) \rangle_H$ $(\forall a, b \in \mathcal{X})$.

Kernel, RKHS

- $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ kernel on $\mathcal{X}$, if
  - $\exists \varphi : \mathcal{X} \to H$(ilbert space) feature map,
  - $k(a, b) = \langle \varphi(a), \varphi(b) \rangle_H$ $(\forall a, b \in \mathcal{X})$.
- Kernel examples: $\mathcal{X} = \mathbb{R}^d$ $(p > 0,\ \theta > 0)$
  - $k(a, b) = (\langle a, b \rangle + \theta)^p$: polynomial,
  - $k(a, b) = e^{-\|a-b\|_2^2/(2\theta^2)}$: Gaussian,
  - $k(a, b) = e^{-\theta\|a-b\|_2}$: Laplacian.

Kernel, RKHS

- $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ kernel on $\mathcal{X}$, if
  - $\exists \varphi : \mathcal{X} \to H$(ilbert space) feature map,
  - $k(a, b) = \langle \varphi(a), \varphi(b) \rangle_H$ ($\forall a, b \in \mathcal{X}$).
- Kernel examples: $\mathcal{X} = \mathbb{R}^d$ ($p > 0$, $\theta > 0$)
  - $k(a, b) = (\langle a, b \rangle + \theta)^p$: polynomial,
  - $k(a, b) = e^{-\|a-b\|_2^2/(2\theta^2)}$: Gaussian,
  - $k(a, b) = e^{-\theta\|a-b\|_2}$: Laplacian.
- In the $H = H(k)$ RKHS ($\exists!$): $\varphi(u) = k(\cdot, u)$.

RKHS: evaluation point of view

- Let $H \subset \mathbb{R}^{\mathcal{X}}$ be a Hilbert space.
- Consider for fixed $x \in \mathcal{X}$ the $\delta_x : f \in H \mapsto f(x) \in \mathbb{R}$ map.
- The evaluation functional is linear:

$$\delta_x(\alpha f + \beta g) = \alpha \delta_x(f) + \beta \delta_x(g).$$

RKHS: evaluation point of view

- Let $H \subset \mathbb{R}^{\mathcal{X}}$ be a Hilbert space.
- Consider for fixed $x \in \mathcal{X}$ the $\delta_x : f \in H \mapsto f(x) \in \mathbb{R}$ map.
- The evaluation functional is linear:

$$\delta_x(\alpha f + \beta g) = \alpha\delta_x(f) + \beta\delta_x(g).$$

- Def.: H is called *RKHS* if δ_x is continuous for $\forall x \in \mathcal{X}$.

RKHS: reproducing point of view

- Let $H \subset \mathbb{R}^{\mathcal{X}}$ be a Hilbert space.
- $k : \mathcal{X} \times \mathcal{X} \to$ is called a *reproducing kernel of H* if for $\forall x \in \mathcal{X}, f \in H$
  1. $k(\cdot, x) \in H$,
  2. $\langle f, k(\cdot, x) \rangle_H = f(x)$ (reproducing property).

RKHS: reproducing point of view

- Let $H \subset \mathbb{R}^{\mathcal{X}}$ be a Hilbert space.
- $k : \mathcal{X} \times \mathcal{X} \to$ is called a *reproducing kernel of H* if for $\forall x \in \mathcal{X}, f \in H$
  1. $k(\cdot, x) \in H$,
  2. $\langle f, k(\cdot, x) \rangle_H = f(x)$ (reproducing property).

  Specifically, $\forall x, y \in \mathcal{X}$

$$k(x, y) = \langle k(\cdot, x), k(\cdot, y) \rangle_H .$$

RKHS: positive-definite point of view

- Let us given a $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ symmetric function.
- k is called *positive definite* if $\forall n \geq 1$, $\forall (a_1, \ldots, a_n) \in \mathbb{R}^n$, $(x_1, \ldots, x_n) \in \mathcal{X}^n$

$$\sum_{i,j=1}^{n} a_i a_j k(x_i, x_j) = \mathbf{a}^T \mathbf{G} \mathbf{a} \geq 0,$$

where $\mathbf{G} = [k(x_i, x_j)]_{i,j=1}^{n}$.

Kernel: example domains ($\mathcal{X}$)

- Euclidean space: $\mathcal{X} = \mathbb{R}^d$.
- Graphs, texts, time series, dynamical systems, distributions.

Kernel: application example – ridge regression

- Given: $\{(x_i, y_i)\}_{i=1}^{\ell}$, $H = H(k)$.
- Task: find $f \in H$ s.t. $f(x_i) \approx y_i$,

$$J(f) = \frac{1}{\ell}\sum_{i=1}^{\ell}[f(x_i) - y_i]^2 + \lambda \|f\|_H^2 \to \min_{f \in H} \quad (\lambda > 0).$$

Kernel: application example – ridge regression

- Given: $\{(x_i, y_i)\}_{i=1}^{\ell}$, $H = H(k)$.
- Task: find $f \in H$ s.t. $f(x_i) \approx y_i$,

$$J(f) = \frac{1}{\ell}\sum_{i=1}^{\ell}[f(x_i) - y_i]^2 + \lambda \|f\|_H^2 \to \min_{f \in H} \quad (\lambda > 0).$$

- Analytical solution, $\mathcal{O}(\ell^3)$ – expensive:

$$f(x) = [k(x_1, x), \ldots, k(x_\ell, x)](\mathbf{G} + \lambda\ell\mathbf{I})^{-1}[y_1; \ldots; y_\ell],$$
$$\mathbf{G} = [k(x_i, x_j)]_{i,j=1}^{\ell}.$$

Kernel: application example – ridge regression

- Given: $\{(x_i, y_i)\}_{i=1}^{\ell}$, $H = H(k)$.
- Task: find $f \in H$ s.t. $f(x_i) \approx y_i$,

$$J(f) = \frac{1}{\ell}\sum_{i=1}^{\ell}[f(x_i) - y_i]^2 + \lambda \|f\|_H^2 \to \min_{f \in H} \quad (\lambda > 0).$$

- Analytical solution, $\mathcal{O}(\ell^3)$ – expensive:

$$f(x) = [k(x_1, x), \ldots, k(x_\ell, x)](\mathbf{G} + \lambda\ell\mathbf{I})^{-1}[y_1; \ldots; y_\ell],$$
$$\mathbf{G} = [k(x_i, x_j)]_{i,j=1}^{\ell}.$$

- **Idea**: $\hat{\mathbf{G}}$, matrix-inversion lemma, fast primal solvers $\to$ RFF.

Focus

- $\mathcal{X} = \mathbb{R}^d$. k: continuous, shift-invariant $[k(\mathbf{x},\mathbf{y}) = \tilde{k}(\mathbf{x}-\mathbf{y})]$.
- By Bochner's theorem:

$$k(\mathbf{x},\mathbf{y}) = \int_{\mathbb{R}^d} e^{i\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y})} \mathrm{d}\Lambda(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} \cos\left(\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y})\right) \mathrm{d}\Lambda(\boldsymbol{\omega}).$$

Focus

- $\mathcal{X} = \mathbb{R}^d$. k: continuous, shift-invariant $[k(\mathbf{x}, \mathbf{y}) = \tilde{k}(\mathbf{x} - \mathbf{y})]$.
- By Bochner's theorem:

$$k(\mathbf{x}, \mathbf{y}) = \int_{\mathbb{R}^d} e^{i\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y})} \mathrm{d}\Lambda(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} \cos\left(\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y})\right) \mathrm{d}\Lambda(\boldsymbol{\omega}).$$

- RFF trick [Rahimi and Recht, 2007] (MC): $\boldsymbol{\omega}_{1:m} := (\boldsymbol{\omega}_j)_{j=1}^m \overset{i.i.d.}{\sim} \Lambda$,

$$\hat{k}(\mathbf{x}, \mathbf{y}) = \frac{1}{m} \sum_{j=1}^m \cos\left(\boldsymbol{\omega}_j^T(\mathbf{x}-\mathbf{y})\right) = \int_{\mathbb{R}^d} \cos\left(\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y})\right) \mathrm{d}\Lambda_m(\boldsymbol{\omega}).$$

RFF – existing guarantee, basically

- Hoeffding inequality + union bound:

$$\left\|k - \hat{k}\right\|_{L^\infty(\mathcal{S})} = \mathcal{O}_p\left(\underbrace{|\mathcal{S}|}_{\text{linear}} \sqrt{\frac{\log m}{m}}\right).$$

RFF – existing guarantee, basically

- Hoeffding inequality + union bound:

$$\left\|k - \hat{k}\right\|_{L^\infty(\mathcal{S})} = \mathcal{O}_p\left(\underbrace{|\mathcal{S}|}_{\text{linear}}\sqrt{\frac{\log m}{m}}\right).$$

- Characteristic function point of view [Csörgő and Totik, 1983] (asymptotic!):
  1. $|\mathcal{S}_m| = e^{o(m)}$ is the optimal rate for a.s. convergence,
  2. For faster growing $|\mathcal{S}_m|$: even convergence in probability fails.

Today: one-page summary

1. Finite-sample L^∞-guarantee $\xrightarrow{\text{specifically}}$

$$\left\|k - \hat{k}\right\|_{L^\infty(\mathcal{S})} = \mathcal{O}_{a.s.}\left(\frac{\sqrt{\log|\mathcal{S}|}}{\sqrt{m}}\right)$$

$\Rightarrow \mathcal{S}$ can grow exponentially $[|\mathcal{S}_m| = e^{o(m)}]$ – optimal!

Today: one-page summary

1. Finite-sample L^∞-guarantee $\xrightarrow{\text{specifically}}$

$$\left\| k - \hat{k} \right\|_{L^\infty(\mathcal{S})} = \mathcal{O}_{a.s.}\left(\frac{\sqrt{\log |\mathcal{S}|}}{\sqrt{m}} \right)$$

$\Rightarrow \mathcal{S}$ can grow exponentially $[|\mathcal{S}_m| = e^{o(m)}]$ – optimal!

2. Finite sample L^p guarantees, $p \in [1, \infty)$.

. . . , where

- Uniform ($p = \infty$), L^p ($1 \le p < \infty$) norm:

$$\|k - \hat{k}\|_{L^\infty(\mathcal{S})} := \sup_{\mathbf{x},\mathbf{y}\in\mathcal{S}} \left| k(\mathbf{x},\mathbf{y}) - \hat{k}(\mathbf{x},\mathbf{y}) \right|,$$

$$\|k - \hat{k}\|_{L^p(\mathcal{S})} := \left(\int_{\mathcal{S}} \int_{\mathcal{S}} |k(\mathbf{x},\mathbf{y}) - \hat{k}(\mathbf{x},\mathbf{y})|^p \,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \right)^{\frac{1}{p}}.$$

Uniform bound: proof idea

1. Empirical process form $[\mathbb{P}g := \int g \mathrm{d}\mathbb{P}]$:

$$\sup_{\mathbf{x},\mathbf{y}\in\mathcal{S}} \left|k(\mathbf{x},\mathbf{y}) - \hat{k}(\mathbf{x},\mathbf{y})\right| = \sup_{g\in\mathcal{G}} |\Lambda g - \Lambda_m g| = \|\Lambda - \Lambda_m\|_{\mathcal{G}}.$$

Uniform bound: proof idea

1. Empirical process form [$\mathbb{P}g := \int g \mathrm{d}\mathbb{P}$]:

$$\sup_{\mathbf{x},\mathbf{y}\in\mathcal{S}} \left| k(\mathbf{x},\mathbf{y}) - \hat{k}(\mathbf{x},\mathbf{y}) \right| = \sup_{g\in\mathcal{G}} |\Lambda g - \Lambda_m g| = \|\Lambda - \Lambda_m\|_{\mathcal{G}}.$$

2. $f(\boldsymbol{\omega}_{1:m}) = \|\Lambda - \Lambda_m\|_{\mathcal{G}}$ concentrates (bounded difference):

$$\|\Lambda - \Lambda_m\|_{\mathcal{G}} \precsim \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \|\Lambda - \Lambda_m\|_{\mathcal{G}} + \frac{1}{\sqrt{m}}.$$

Uniform bound: proof idea

1. Empirical process form [$\mathbb{P}g := \int g \mathrm{d}\mathbb{P}$]:

$$\sup_{\mathbf{x},\mathbf{y}\in\mathcal{S}} \left| k(\mathbf{x},\mathbf{y}) - \hat{k}(\mathbf{x},\mathbf{y}) \right| = \sup_{g\in\mathcal{G}} |\Lambda g - \Lambda_m g| = \|\Lambda - \Lambda_m\|_{\mathcal{G}}.$$

2. $f(\omega_{1:m}) = \|\Lambda - \Lambda_m\|_{\mathcal{G}}$ concentrates (bounded difference):

$$\|\Lambda - \Lambda_m\|_{\mathcal{G}} \lesssim \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \|\Lambda - \Lambda_m\|_{\mathcal{G}} + \frac{1}{\sqrt{m}}.$$

3. $\mathcal{G}$ is 'nice' (uniformly bounded, separable Carathéodory) $\Rightarrow$

$$\mathbb{E}_{\boldsymbol{\omega}_{1:m}} \|\Lambda - \Lambda_m\|_{\mathcal{G}} \lesssim \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \underbrace{\mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right)}_{\mathbb{E}_{\epsilon} \sup_{g\in\mathcal{G}} \left| \frac{1}{m} \sum_{j=1}^{m} \epsilon_j g(\omega_j) \right|}.$$

Uniform bound: proof idea – continued

4. Using Dudley's entropy bound:

$$\mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) \lesssim \frac{1}{\sqrt{m}} \int_0^{|\mathcal{G}|_{L^2(\Lambda_m)}} \sqrt{\log \mathcal{N}(\mathcal{G}, L^2(\Lambda_m), u)} \mathrm{d}u.$$

Uniform bound: proof idea – continued

4. Using Dudley's entropy bound:

$$\mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) \lesssim \frac{1}{\sqrt{m}} \int_0^{|\mathcal{G}|_{L^2(\Lambda_m)}} \sqrt{\log \mathcal{N}(\mathcal{G}, L^2(\Lambda_m), u)} \mathrm{d}u.$$

5. $\mathcal{G}$ is smoothly parameterized by a compact set $\Rightarrow$

$$\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), u) \leq \left(\frac{4|\mathcal{S}|A}{u} + 1\right)^d, \; A(\boldsymbol{\omega}_{1:m}) = \sqrt{\frac{1}{m}\sum_{j=1}^m \|\boldsymbol{\omega}_j\|_2^2}.$$

Uniform bound: proof idea – continued

4. Using Dudley's entropy bound:

$$\mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) \lesssim \frac{1}{\sqrt{m}} \int_0^{|\mathcal{G}|_{L^2(\Lambda_m)}} \sqrt{\log \mathcal{N}(\mathcal{G}, L^2(\Lambda_m), u)} \mathrm{d}u.$$

5. $\mathcal{G}$ is smoothly parameterized by a compact set $\Rightarrow$

$$\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), u) \leq \left(\frac{4|\mathcal{S}|A}{u} + 1\right)^d, \; A(\boldsymbol{\omega}_{1:m}) = \sqrt{\frac{1}{m}\sum_{j=1}^m \|\boldsymbol{\omega}_j\|_2^2}.$$

6. Putting together [$|\mathcal{G}|_{L^2(\Lambda_m)} \leq 2$, Jensen inequality] we get the result.

Step-1: empirical process form

- Recall the notation:

$$\Lambda g = \int_{\mathbb{R}^d} g(\omega)\mathrm{d}\Lambda(\omega), \quad \Lambda_m g = \int_{\mathbb{R}^d} g(\omega)\mathrm{d}\Lambda_m(\omega) = \frac{1}{m}\sum_{j=1}^{m} g(\omega_j).$$

Step-1: empirical process form

- Recall the notation:

$$\Lambda g = \int_{\mathbb{R}^d} g(\boldsymbol{\omega}) \mathrm{d}\Lambda(\boldsymbol{\omega}), \quad \Lambda_m g = \int_{\mathbb{R}^d} g(\boldsymbol{\omega}) \mathrm{d}\Lambda_m(\boldsymbol{\omega}) = \frac{1}{m} \sum_{j=1}^{m} g(\boldsymbol{\omega}_j).$$

- Reformulation of the objective:

$$\sup_{\mathbf{x},\mathbf{y} \in \mathcal{S}} \left| k(\mathbf{x}, \mathbf{y}) - \hat{k}(\mathbf{x}, \mathbf{y}) \right| = \sup_{g \in \mathcal{G}} |\Lambda g - \Lambda_m g| =: \|\Lambda - \Lambda_m\|_{\mathcal{G}} ,$$

where

$$\mathcal{G} = \{g_{\mathbf{z}} : \mathbf{z} \in \mathcal{S}_\Delta\},$$
$$\mathcal{S}_\Delta = \mathcal{S} - \mathcal{S} = \{\mathbf{x} - \mathbf{y} : \mathbf{x}, \mathbf{y} \in \mathcal{S}\},$$
$$g_{\mathbf{z}} : \boldsymbol{\omega} \in \mathbb{R}^d \mapsto \cos\left(\boldsymbol{\omega}^T \mathbf{z}\right) \in \mathbb{R}.$$

Step-2: bounded diff. property of $f(\omega_{1:m}) = \|\Lambda - \Lambda_m\|_{\mathcal{G}}$

McDiarmid inequality: Let $\omega_1, \ldots, \omega_m \in D$ be independent r.v.-s, and $f : D^m \to \mathbb{R}$ satisfy the bounded diff. property ($\forall r$):

$$\sup_{u_1,\ldots,u_m,u_r' \in D} \left| f(u_1,\ldots,u_m) - f(u_1,\ldots,u_{r-1},u_r',u_{r+1},\ldots,u_m) \right| \leq c_r.$$

Then for $\forall \beta > 0$

$$\mathbb{P}\left(f(\omega_{1:m}) - \mathbb{E}\left[f(\omega_{1:m})\right] \geq \beta\right) \leq e^{-\frac{2\beta^2}{\sum_{r=1}^m c_r^2}}.$$

Step-2: bounded difference property of $\|\Lambda - \Lambda_m\|_{\mathcal{G}}$

Our choice: $f(\omega_1, \ldots, \omega_m) := \|\Lambda - \Lambda_m\|_{\mathcal{G}}$.

$$|f(\omega_1, \ldots, \omega_{r-1}, \omega_r, \omega_{r+1}, \ldots, \omega_m) - f(\omega_1, \ldots, \omega_{r-1}, \omega_r', \omega_{r+1}, \ldots, \omega_m)| =$$

$$= \left| \sup_{g \in \mathcal{G}} \left| \Lambda g - \frac{1}{m} \sum_{j=1} g(\omega_j) \right| - \sup_{g \in \mathcal{G}} \left| \Lambda g - \frac{1}{m} \sum_{j=1} g(\omega_j) + \frac{1}{m} \left[g(\omega_r) - g(\omega_r') \right] \right| \right|$$

Step-2: bounded difference property of $\|\Lambda - \Lambda_m\|_{\mathcal{G}}$

Our choice: $f(\omega_1, \ldots, \omega_m) := \|\Lambda - \Lambda_m\|_{\mathcal{G}}$.

$$|f(\omega_1, \ldots, \omega_{r-1}, \omega_r, \omega_{r+1}, \ldots, \omega_m) - f(\omega_1, \ldots, \omega_{r-1}, \omega_r', \omega_{r+1}, \ldots, \omega_m)| =$$

$$= \left| \sup_{g \in \mathcal{G}} \left| \Lambda g - \frac{1}{m} \sum_{j=1} g(\omega_j) \right| - \sup_{g \in \mathcal{G}} \left| \Lambda g - \frac{1}{m} \sum_{j=1} g(\omega_j) + \frac{1}{m} \left[g(\omega_r) - g(\omega_r') \right] \right| \right|$$

$$\overset{(*)}{\leq} \frac{1}{m} \sup_{g \in \mathcal{G}} \left| g(\omega_r) - g(\omega_r') \right|$$

Step-2: bounded difference property of $\|\Lambda - \Lambda_m\|_{\mathcal{G}}$

Our choice: $f(\omega_1, \ldots, \omega_m) := \|\Lambda - \Lambda_m\|_{\mathcal{G}}$.

$$|f(\omega_1, \ldots, \omega_{r-1}, \omega_r, \omega_{r+1}, \ldots, \omega_m) - f(\omega_1, \ldots, \omega_{r-1}, \omega_r', \omega_{r+1}, \ldots, \omega_m)| =$$

$$= \left| \sup_{g \in \mathcal{G}} \left| \Lambda g - \frac{1}{m} \sum_{j=1} g(\omega_j) \right| - \sup_{g \in \mathcal{G}} \left| \Lambda g - \frac{1}{m} \sum_{j=1} g(\omega_j) + \frac{1}{m} \left[g(\omega_r) - g(\omega_r') \right] \right| \right|$$

$$\overset{(*)}{\leq} \frac{1}{m} \sup_{g \in \mathcal{G}} \left| g(\omega_r) - g(\omega_r') \right| \leq \frac{1}{m} \sup_{g \in \mathcal{G}} \left(\left| g(\omega_r) \right| + \left| g(\omega_r') \right| \right)$$

Step-2: bounded difference property of $\|\Lambda - \Lambda_m\|_{\mathcal{G}}$

Our choice: $f(\omega_1, \ldots, \omega_m) := \|\Lambda - \Lambda_m\|_{\mathcal{G}}$.

$$|f(\omega_1, \ldots, \omega_{r-1}, \omega_r, \omega_{r+1}, \ldots, \omega_m) - f(\omega_1, \ldots, \omega_{r-1}, \omega_r', \omega_{r+1}, \ldots, \omega_m)| =$$

$$= \left| \sup_{g \in \mathcal{G}} \left| \Lambda g - \frac{1}{m} \sum_{j=1} g(\omega_j) \right| - \sup_{g \in \mathcal{G}} \left| \Lambda g - \frac{1}{m} \sum_{j=1} g(\omega_j) + \frac{1}{m} \left[g(\omega_r) - g(\omega_r') \right] \right| \right|$$

$$\overset{(*)}{\leq} \frac{1}{m} \sup_{g \in \mathcal{G}} |g(\omega_r) - g(\omega_r')| \leq \frac{1}{m} \sup_{g \in \mathcal{G}} \left(|g(\omega_r)| + |g(\omega_r')| \right)$$

$$\leq \frac{1}{m} \left[\sup_{g \in \mathcal{G}} |g(\omega_r)| + \sup_{g \in \mathcal{G}} |g(\omega_r')| \right]$$

Step-2: bounded difference property of $\|\Lambda - \Lambda_m\|_{\mathcal{G}}$

Our choice: $f(\omega_1, \ldots, \omega_m) := \|\Lambda - \Lambda_m\|_{\mathcal{G}}$.

$$|f(\omega_1, \ldots, \omega_{r-1}, \omega_r, \omega_{r+1}, \ldots, \omega_m) - f(\omega_1, \ldots, \omega_{r-1}, \omega_r', \omega_{r+1}, \ldots, \omega_m)| =$$

$$= \left| \sup_{g \in \mathcal{G}} \left| \Lambda g - \frac{1}{m} \sum_{j=1} g(\omega_j) \right| - \sup_{g \in \mathcal{G}} \left| \Lambda g - \frac{1}{m} \sum_{j=1} g(\omega_j) + \frac{1}{m} \left[g(\omega_r) - g(\omega_r') \right] \right| \right|$$

$$\overset{(*)}{\leq} \frac{1}{m} \sup_{g \in \mathcal{G}} \left| g(\omega_r) - g(\omega_r') \right| \leq \frac{1}{m} \sup_{g \in \mathcal{G}} \left(|g(\omega_r)| + |g(\omega_r')| \right)$$

$$\leq \frac{1}{m} \left[\sup_{g \in \mathcal{G}} |g(\omega_r)| + \sup_{g \in \mathcal{G}} |g(\omega_r')| \right] \leq \frac{1+1}{m} = \frac{2}{m},$$

$(*)$: next slide.

Step-2: (*) = reverse triangle inequality with sup

- Lemma: $\mathcal{G}$: set of functions, $a, b : \mathcal{G} \to \mathbb{R}$ maps; then

$$\left| \sup_{g \in \mathcal{G}} |a(g)| - \sup_{g \in \mathcal{G}} |a(g) + b(g)| \right|$$

Step-2: (*) = reverse triangle inequality with sup

- Lemma: $\mathcal{G}$: set of functions, $a, b : \mathcal{G} \to \mathbb{R}$ maps; then

$$\left| \sup_{g\in\mathcal{G}} |a(g)| - \sup_{g\in\mathcal{G}} |a(g) + b(g)| \right| \leq \sup_{g\in\mathcal{G}} |b(g)|.$$

Step-2: (*) = reverse triangle inequality with sup

- Lemma: $\mathcal{G}$: set of functions, $a, b : \mathcal{G} \to \mathbb{R}$ maps; then

$$\left| \sup_{g\in\mathcal{G}} |a(g)| - \sup_{g\in\mathcal{G}} |a(g) + b(g)| \right| \leq \sup_{g\in\mathcal{G}} |b(g)|.$$

- Proof: combine

$$sup_{g\in\mathcal{G}} |a(g) + b(g)| \leq \sup_{g\in\mathcal{G}} \left(|a(g)| + |b(g)|\right) \leq \sup_{g\in\mathcal{G}} |a(g)| + \sup_{g\in\mathcal{G}} |b(g)|,$$

Step-2: (*) = reverse triangle inequality with sup

- Lemma: $\mathcal{G}$: set of functions, $a, b : \mathcal{G} \to \mathbb{R}$ maps; then

$$\left| \sup_{g\in\mathcal{G}} |a(g)| - \sup_{g\in\mathcal{G}} |a(g) + b(g)| \right| \le \sup_{g\in\mathcal{G}} |b(g)|.$$

- Proof: combine

$$\sup_{g\in\mathcal{G}} |a(g) + b(g)| \le \sup_{g\in\mathcal{G}} (|a(g)| + |b(g)|) \le \sup_{g\in\mathcal{G}} |a(g)| + \sup_{g\in\mathcal{G}} |b(g)|,$$

$$\begin{aligned} \sup_{g\in\mathcal{G}} |a(g)| &= \sup_{g\in\mathcal{G}} |a(g) + b(g) - b(g)| \\ &\le \sup_{g\in\mathcal{G}} |a(g) + b(g)| + \sup_{g\in\mathcal{G}} |b(g)|. \end{aligned}$$

Step-2: (*) = reverse triangle inequality with sup

- Lemma: $\mathcal{G}$: set of functions, $a, b : \mathcal{G} \to \mathbb{R}$ maps; then

$$\left| \sup_{g\in\mathcal{G}} |a(g)| - \sup_{g\in\mathcal{G}} |a(g) + b(g)| \right| \le \sup_{g\in\mathcal{G}} |b(g)|.$$

- Proof: combine

$$sup_{g\in\mathcal{G}} |a(g) + b(g)| \le \sup_{g\in\mathcal{G}} (|a(g)| + |b(g)|) \le \sup_{g\in\mathcal{G}} |a(g)| + \sup_{g\in\mathcal{G}} |b(g)|,$$

$$\sup_{g\in\mathcal{G}} |a(g)| = \sup_{g\in\mathcal{G}} |a(g) + b(g) - b(g)|$$

$$\le \sup_{g\in\mathcal{G}} |a(g) + b(g)| + \sup_{g\in\mathcal{G}} |b(g)|.$$

$$\Rightarrow \pm \left[\sup_{g\in\mathcal{G}} |a(g)| - \sup_{g\in\mathcal{G}} |a(g) + b(g)| \right] \le \sup_{g\in\mathcal{G}} |b(g)|.$$

- Our choice: $a(g) = \Lambda g - \frac{1}{m} \sum_{j=1} g(\omega_j)$, $b(g) = \frac{1}{m} [g(\omega_r) - g(\omega'_r)]$.

Step-2

Applying McDiarmid to f $(c_r = \frac{2}{m})$: for $\forall \tau > 0$ with probability $1 - e^{-\tau}$

$$\|\Lambda - \Lambda_m\|_{\mathcal{G}} \leq \underbrace{\mathbb{E}_{\boldsymbol{\omega}_{1:m}} \|\Lambda - \Lambda_m\|_{\mathcal{G}}}_{\text{Step-3: bounding this term}} + \frac{\sqrt{2\tau}}{\sqrt{m}}.$$

Step-3: bounding $\mathbb{E}_{\omega_{1:m}} \|\Lambda - \Lambda_m\|_{\mathcal{G}}$

$\mathcal{G} = \{g_{\mathbf{z}} : \mathbf{z} \in \mathcal{S}_\Delta\}$ is a separable Carathéodory family, i.e.

1. $\boldsymbol{\omega} \mapsto \cos\left(\boldsymbol{\omega}^T \mathbf{z}\right)$: measurable for $\forall \mathbf{z} \in \mathcal{S}_\Delta$.

Step-3: bounding $\mathbb{E}_{\omega_{1:m}} \|\Lambda - \Lambda_m\|_{\mathcal{G}}$

$\mathcal{G} = \{g_{\mathbf{z}} : \mathbf{z} \in \mathcal{S}_\Delta\}$ is a separable Carathéodory family, i.e.

1. $\boldsymbol{\omega} \mapsto \cos\left(\boldsymbol{\omega}^T \mathbf{z}\right)$: measurable for $\forall \mathbf{z} \in \mathcal{S}_\Delta$.
2. $\mathbf{z} \mapsto \cos\left(\boldsymbol{\omega}^T \mathbf{z}\right)$: continuous for $\forall \boldsymbol{\omega}$.

Step-3: bounding $\mathbb{E}_{\omega_{1:m}} \|\Lambda - \Lambda_m\|_{\mathcal{G}}$

$\mathcal{G} = \{g_{\mathbf{z}} : \mathbf{z} \in \mathcal{S}_{\Delta}\}$ is a separable Carathéodory family, i.e.

1. $\boldsymbol{\omega} \mapsto \cos\left(\boldsymbol{\omega}^T \mathbf{z}\right)$: measurable for $\forall \mathbf{z} \in \mathcal{S}_{\Delta}$.
2. $\mathbf{z} \mapsto \cos\left(\boldsymbol{\omega}^T \mathbf{z}\right)$: continuous for $\forall \boldsymbol{\omega}$.
3. $\mathbb{R}^d$ is separable, $\mathcal{S}_{\Delta} \subseteq \mathbb{R}^d \Rightarrow \mathcal{S}_{\Delta}$: separable.

Step-3: bounding $\mathbb{E}_{\omega_{1:m}} \|\Lambda - \Lambda_m\|_{\mathcal{G}}$

$\mathcal{G} = \{g_{\mathbf{z}} : \mathbf{z} \in \mathcal{S}_\Delta\}$ is a separable Carathéodory family, i.e.

1. $\boldsymbol{\omega} \mapsto \cos\left(\boldsymbol{\omega}^T \mathbf{z}\right)$: measurable for $\forall \mathbf{z} \in \mathcal{S}_\Delta$.
2. $\mathbf{z} \mapsto \cos\left(\boldsymbol{\omega}^T \mathbf{z}\right)$: continuous for $\forall \boldsymbol{\omega}$.
3. $\mathbb{R}^d$ is separable, $\mathcal{S}_\Delta \subseteq \mathbb{R}^d \Rightarrow \mathcal{S}_\Delta$: separable.

Thus, by [Steinwart and Christmann, 2008, Prop. 7.10]

$$\mathbb{E}_{\boldsymbol{\omega}_{1:m}} \|\Lambda - \Lambda_m\|_{\mathcal{G}} \le 2\mathbb{E}_{\boldsymbol{\omega}_{1:m}} [\underbrace{\mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right)}_{:=\mathbb{E}_\epsilon \sup_{g\in\mathcal{G}} \left| \frac{1}{m} \sum_{j=1}^m \epsilon_j g(\omega_j) \right|}]$$

using the uniformly boundedness of $\mathcal{G}$ $\left(\sup_{g\in\mathcal{G}} \|g\|_\infty \le 1 < \infty\right)$.

Step-4: bounding $\mathcal{R}$

Using Dudley's entropy integral [Bousquet, 2003, Eq. (4.4)]:

$$\mathcal{R}\left(\mathcal{G}, (\boldsymbol{\omega}_j)_{j=1}^m\right) \le \frac{8\sqrt{2}}{\sqrt{m}} \int_0^{|\mathcal{G}|_{L^2(\Lambda_m)}} \sqrt{\log \mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)}\mathrm{d}r,$$

Step-4: bounding $\mathcal{R}$

Using Dudley's entropy integral [Bousquet, 2003, Eq. (4.4)]:

$$\mathcal{R}\left(\mathcal{G},(\boldsymbol{\omega}_j)_{j=1}^m\right) \le \frac{8\sqrt{2}}{\sqrt{m}} \int_0^{|\mathcal{G}|_{L^2(\Lambda_m)}} \sqrt{\log \mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)}\mathrm{d}r,$$

where

- $L^2(\Lambda_m) = L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \Lambda_m)$, $\|g\|_{L^2(\Lambda_m)} = \sqrt{\frac{1}{m}\sum_{j=1}^m g^2(\omega_j)}$,

Step-4: bounding $\mathcal{R}$

Using Dudley's entropy integral [Bousquet, 2003, Eq. (4.4)]:

$$\mathcal{R}\left(\mathcal{G},(\boldsymbol{\omega}_j)_{j=1}^m\right) \le \frac{8\sqrt{2}}{\sqrt{m}} \int_0^{|\mathcal{G}|_{L^2(\Lambda_m)}} \sqrt{\log \mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)}\mathrm{d}r,$$

where

- $L^2(\Lambda_m) = L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \Lambda_m)$, $\|g\|_{L^2(\Lambda_m)} = \sqrt{\frac{1}{m}\sum_{j=1}^m g^2(\omega_j)}$,
- $|\mathcal{G}|_{L^2(\Lambda_m)} = \sup_{g_1,g_2\in\mathcal{G}} \|g_1 - g_2\|_{L^2(\Lambda_m)}$,

Step-4: bounding $\mathcal{R}$

Using Dudley's entropy integral [Bousquet, 2003, Eq. (4.4)]:

$$\mathcal{R}\left(\mathcal{G}, (\omega_j)_{j=1}^m\right) \le \frac{8\sqrt{2}}{\sqrt{m}} \int_0^{|\mathcal{G}|_{L^2(\Lambda_m)}} \sqrt{\log \mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)}\mathrm{d}r,$$

where

- $L^2(\Lambda_m) = L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \Lambda_m)$, $\|g\|_{L^2(\Lambda_m)} = \sqrt{\frac{1}{m}\sum_{j=1}^m g^2(\omega_j)}$,
- $|\mathcal{G}|_{L^2(\Lambda_m)} = \sup_{g_1,g_2\in\mathcal{G}} \|g_1 - g_2\|_{L^2(\Lambda_m)}$,
- $\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)$: r-covering number.
  - r-net: $S \subseteq \mathcal{G}$, for $\forall g \in \mathcal{G}$ $\exists s \in S$ such that $\|g - s\|_{L^2(\Lambda_m)} \le r$.
  - $\mathcal{N}$: size of the smallest r-net of $\mathcal{G}$.

Step-5: bound on $|\mathcal{G}|_{L^2(\Lambda_m)}$

$$|\mathcal{G}|_{L^2(\Lambda_m)} = \sup_{g_1, g_2 \in \mathcal{G}} \|g_1 - g_2\|_{L^2(\Lambda_m)} \leq \sup_{g_1, g_2 \in \mathcal{G}} \left(\|g_1\|_{L^2(\Lambda_m)} + \|g_2\|_{L^2(\Lambda_m)} \right)$$

$$\leq \sup_{g_1 \in \mathcal{G}} \|g_1\|_{L^2(\Lambda_m)} + \sup_{g_1 \in \mathcal{G}} \|g_2\|_{L^2(\Lambda_m)} \overset{(*)}{\leq} 2,$$

Step-5: bound on $|\mathcal{G}|_{L^2(\Lambda_m)}$

$$|\mathcal{G}|_{L^2(\Lambda_m)} = \sup_{g_1, g_2 \in \mathcal{G}} \|g_1 - g_2\|_{L^2(\Lambda_m)} \leq \sup_{g_1, g_2 \in \mathcal{G}} \left(\|g_1\|_{L^2(\Lambda_m)} + \|g_2\|_{L^2(\Lambda_m)} \right)$$

$$\leq \sup_{g_1 \in \mathcal{G}} \|g_1\|_{L^2(\Lambda_m)} + \sup_{g_1 \in \mathcal{G}} \|g_2\|_{L^2(\Lambda_m)} \overset{(*)}{\leq} 2,$$

$$\sup_{g \in \mathcal{G}} \|g\|_{L^2(\Lambda_m)} \overset{(*)}{=} \sup_{\mathbf{z} \in \mathcal{S}_\Delta} \sqrt{\frac{1}{m} \sum_{j=1}^{m} g_{\mathbf{z}}^2(\boldsymbol{\omega}_j)} = \sup_{\mathbf{z} \in \mathcal{S}_\Delta} \sqrt{\frac{1}{m} \sum_{j=1}^{m} \left[\cos\left(\boldsymbol{\omega}_j^T \mathbf{z}\right)\right]^2} \leq 1.$$

Step-5: bound on $\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)$

Let $g_{\mathbf{z}_1}, g_{\mathbf{z}_2} \in \mathcal{G}$. We want to bound $\|g_{\mathbf{z}_1} - g_{\mathbf{z}_2}\|_{L^2(\Lambda_m)}$. One term:

$$\left|\cos\left(\boldsymbol{\omega}^T \mathbf{z}_1\right) - \cos\left(\boldsymbol{\omega}^T \mathbf{z}_2\right)\right| \overset{(*)}{\leq} \left\|\nabla_{\mathbf{z}} \cos\left(\boldsymbol{\omega}^T \mathbf{z}_c\right)\right\|_2 \|\mathbf{z}_1 - \mathbf{z}_2\|_2$$
$$= \left\|-\sin\left(\boldsymbol{\omega}^T \mathbf{z}_c\right)\boldsymbol{\omega}\right\|_2 \|\mathbf{z}_1 - \mathbf{z}_2\|_2$$
$$\leq \|\boldsymbol{\omega}\|_2 \|\mathbf{z}_1 - \mathbf{z}_2\|_2,$$

where $(*)$: mean-value theorem, $\mathbf{z}_c \in (\mathbf{z}_1, \mathbf{z}_2)$.

Step-5: bound on $\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)$

- Smooth parameterization:

$$\left\|g_{\mathbf{z}_1} - g_{\mathbf{z}_2}\right\|_{L^2(\Lambda_m)} \leq \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(\left\|\boldsymbol{\omega}_j\right\|_2 \left\|\mathbf{z}_1 - \mathbf{z}_2\right\|_2\right)^2}$$

$$= \left\|\mathbf{z}_1 - \mathbf{z}_2\right\|_2 \underbrace{\sqrt{\frac{1}{m}\sum_{j=1}^{m}\left\|\boldsymbol{\omega}_j\right\|_2^2}}_{=:A=A(\boldsymbol{\omega}_{1:m})}.$$

- r-net on $(\mathcal{S}_\Delta, \|\cdot\|_2) \Rightarrow r' = rA$-net on $(\mathcal{G}, L^2(\Lambda_m))$.

Step-5: bound on $\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)$

- Until now:

$$\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r) \le \mathcal{N}\left(\mathcal{S}_\Delta, \|\cdot\|_2, \frac{r}{A}\right).$$

Step-5: bound on $\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)$

- Until now:

$$\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r) \le \mathcal{N}\left(\mathcal{S}_\Delta, \|\cdot\|_2, \frac{r}{A}\right).$$

- $\mathcal{S}_\Delta$ can be covered by a ball of radius $\frac{|\mathcal{S}_\Delta|}{2} \le \frac{2|\mathcal{S}|}{2} = |\mathcal{S}|$.

Step-5: bound on $\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)$

- Until now:

$$\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r) \leq \mathcal{N}\left(\mathcal{S}_\Delta, \|\cdot\|_2, \frac{r}{A}\right).$$

- $\mathcal{S}_\Delta$ can be covered by a ball of radius $\frac{|\mathcal{S}_\Delta|}{2} \leq \frac{2|\mathcal{S}|}{2} = |\mathcal{S}|$.
- ϵ-covering number of $\|\cdot\|_2$-ball in $\mathbb{R}^d$ with radius R: $\leq (\frac{4R}{\epsilon} + 1)^d$ [van de Geer, 2009, Lemma 2.5] $\Rightarrow$

Step-5: bound on $\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)$

- Until now:

$$\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r) \le \mathcal{N}\left(\mathcal{S}_\Delta, \|\cdot\|_2, \frac{r}{A}\right).$$

- $\mathcal{S}_\Delta$ can be covered by a ball of radius $\frac{|\mathcal{S}_\Delta|}{2} \le \frac{2|\mathcal{S}|}{2} = |\mathcal{S}|$.
- ϵ-covering number of $\|\cdot\|_2$-ball in $\mathbb{R}^d$ with radius R: $\le (\frac{4R}{\epsilon} + 1)^d$ [van de Geer, 2009, Lemma 2.5] $\Rightarrow$
- We got [$\epsilon = \frac{r}{A}$, $R = |\mathcal{S}|$]:

$$\mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r) \le \left(\frac{4|\mathcal{S}|A}{r} + 1\right)^d.$$

Step-6: Putting together

$$\mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) \le \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^2 \sqrt{\log\left(\frac{4|\mathcal{S}|A}{r} + 1\right)} \,\mathrm{d}r$$

Step-6: Putting together

$$\mathcal{R}\left(\mathcal{G}, \omega_{1:m}\right) \le \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^2 \sqrt{\log\left(\frac{4|\mathcal{S}|A}{r} + 1\right)} \mathrm{d}r$$

$$\stackrel{(a)}{\le} \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^2 \sqrt{\log\left(\frac{4|\mathcal{S}|A + 2}{r}\right)} \mathrm{d}r$$

(a): $r \le 2$

Step-6: Putting together

$$\mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) \le \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^2 \sqrt{\log\left(\frac{4|\mathcal{S}|A}{r} + 1\right)} \mathrm{d}r$$

$$\overset{(a)}{\le} \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^2 \sqrt{\log\left(\frac{4|\mathcal{S}|A + 2}{r}\right)} \mathrm{d}r$$

(a): $r \le 2$, (b): $2|\mathcal{S}|A + 1 \le (2|\mathcal{S}| + 1)(A + 1)$.

Step-6: Putting together

$$\mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) \le \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^2 \sqrt{\log\left(\frac{4|\mathcal{S}|A}{r} + 1\right)} \,\mathrm{d}r$$
$$\stackrel{(a)}{\le} \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^2 \sqrt{\log\left(\frac{4|\mathcal{S}|A + 2}{r}\right)} \,\mathrm{d}r$$
$$\stackrel{(b)}{\le} \frac{8\sqrt{2d}}{\sqrt{m}} \left[\int_0^2 \sqrt{\log\frac{2\left(2|\mathcal{S}|+1\right)}{r}} \,\mathrm{d}r + 2\sqrt{\log(A+1)}\right].$$

(a): $r \le 2$, (b): $2|\mathcal{S}|A + 1 \le \left(2|\mathcal{S}|+1\right)\left(A+1\right)$.

Step 6: Putting together

$$\mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) \stackrel{(a)}{\leq} \frac{16\sqrt{2d}}{\sqrt{m}}\left(\int_0^1 \sqrt{\log \frac{B+1}{r}}\,\mathrm{d}r + \sqrt{\log(A+1)}\right),$$

(a): change of variables, $B := 2|\mathcal{S}|$

Step 6: Putting together

$$\mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) \overset{(a)}{\leq} \frac{16\sqrt{2d}}{\sqrt{m}} \left(\int_0^1 \sqrt{\log \frac{B+1}{r}} \, \mathrm{d}r + \sqrt{\log(A+1)} \right),$$

(a): change of variables, $B := 2|\mathcal{S}|$, (b): $\int_0^1 \sqrt{\log \frac{a}{\epsilon}} \, \mathrm{d}\epsilon \leq \sqrt{\log a} + \frac{1}{2\sqrt{\log a}}$ $(a > 1)$.

Step 6: Putting together

$$\mathcal{R}(\mathcal{G}, \boldsymbol{\omega}_{1:m}) \overset{(a)}{\leq} \frac{16\sqrt{2d}}{\sqrt{m}} \left(\int_0^1 \sqrt{\log \frac{B+1}{r}} \, \mathrm{d}r + \sqrt{\log(A+1)} \right),$$

$$\overset{(b)}{\leq} \frac{16\sqrt{2d}}{\sqrt{m}} \left[\sqrt{\log(B+1)} + \frac{1}{2\sqrt{\log(B+1)}} + \sqrt{\log(A+1)} \right],$$

(a): change of variables, $B := 2|\mathcal{S}|$, (b):
$\int_0^1 \sqrt{\log \frac{a}{\epsilon}} \, \mathrm{d}\epsilon \leq \sqrt{\log a} + \frac{1}{2\sqrt{\log a}}$ $(a > 1)$. $\Rightarrow$ by Jensen

Uniform guarantee

Let k be continuous, $\sigma^2 := \int_{\mathbb{R}^d} \|\omega\|^2 \, d\Lambda(\omega) < \infty$. Then for $\forall \tau > 0$ and compact set $\mathcal{S} \subset \mathbb{R}^d$

$$\Lambda^m \left(\|\hat{k} - k\|_{L^\infty(\mathcal{S})} \geq \frac{h(d, |\mathcal{S}|, \sigma) + \sqrt{2\tau}}{\sqrt{m}} \right) \leq e^{-\tau},$$

$$h(d, |\mathcal{S}|, \sigma) := 32\sqrt{2d \log(2|\mathcal{S}| + 1)} + 16\sqrt{\frac{2d}{\log(2|\mathcal{S}| + 1)}} + 32\sqrt{2d \log(\sigma + 1)}.$$

Consequence-1 (Borel-Cantelli lemma)

- A.s. convergence on compact sets: $\hat{k} \xrightarrow{m\to\infty} k$ at rate $\sqrt{\frac{\log|\mathcal{S}|}{m}}$.

Consequence-1 (Borel-Cantelli lemma)

- A.s. convergence on compact sets: $\hat{k} \xrightarrow{m\to\infty} k$ at rate $\sqrt{\frac{\log|\mathcal{S}|}{m}}$.
- Growing diameter:
    - $\frac{\log|\mathcal{S}_m|}{m} \xrightarrow{m\to\infty} 0$ is enough (i.e., $|\mathcal{S}_m| = e^{o(m)}$).

Consequence-1 (Borel-Cantelli lemma)

- A.s. convergence on compact sets: $\hat{k} \xrightarrow{m\to\infty} k$ at rate $\sqrt{\frac{\log|\mathcal{S}|}{m}}$.
- Growing diameter:
    - $\frac{\log|\mathcal{S}_m|}{m} \xrightarrow{m\to\infty} 0$ is enough (i.e., $|\mathcal{S}_m| = e^{o(m)}$).
- Specifically:
    - *asymptotic* optimality [Csörgő and Totik, 1983, Theorem 2] (if $k(\mathbf{z})$ vanishes at ∞).

Consequence-2: L^p guarantee $(1 \le p)$

Idea:

- Note that

$$\|\hat{k}-k\|_{L^p(\mathcal{S})} := \left(\int_{\mathcal{S}}\int_{\mathcal{S}} |\hat{k}(\mathbf{x},\mathbf{y})-k(\mathbf{x},\mathbf{y})|^p \,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}\right)^{\frac{1}{p}} \le \|\hat{k}-k\|_{L^\infty(\mathcal{S})} \mathrm{vol}^{2/p}(\mathcal{S}).$$

Consequence-2: L^p guarantee $(1 \le p)$

Idea:

- Note that
$$\|\hat{k} - k\|_{L^p(\mathcal{S})} := \left(\int_{\mathcal{S}} \int_{\mathcal{S}} |\hat{k}(\mathbf{x},\mathbf{y}) - k(\mathbf{x},\mathbf{y})|^p \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{y} \right)^{\frac{1}{p}} \le \|\hat{k} - k\|_{L^\infty(\mathcal{S})} \mathrm{vol}^{2/p}(\mathcal{S}).$$
- $\mathrm{vol}(\mathcal{S}) \le \mathrm{vol}(B)$, where $B := \left\{ \mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \le \frac{|\mathcal{S}|}{2} \right\}$,

Consequence-2: L^p guarantee $(1 \le p)$

Idea:

- Note that

$$\|\hat{k} - k\|_{L^p(\mathcal{S})} := \left(\int_{\mathcal{S}} \int_{\mathcal{S}} |\hat{k}(\mathbf{x}, \mathbf{y}) - k(\mathbf{x}, \mathbf{y})|^p \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{y} \right)^{\frac{1}{p}} \le \|\hat{k} - k\|_{L^\infty(\mathcal{S})} \mathrm{vol}^{2/p}(\mathcal{S}).$$

- $\mathrm{vol}(\mathcal{S}) \le \mathrm{vol}(B)$, where $B := \left\{ \mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \le \frac{|\mathcal{S}|}{2} \right\}$,
- $\mathrm{vol}(B) = \frac{\pi^{d/2}|\mathcal{S}|^d}{2^d \Gamma\left(\frac{d}{2}+1\right)}$, $\Gamma(t) = \int_0^\infty u^{t-1} e^{-u} \, \mathrm{d}u$. $\Rightarrow$

L^p guarantee

Under the previous assumptions, and $1 \leq p < \infty$:

$$\Lambda^m \left(\|\hat{k} - k\|_{L^p(\mathcal{S})} \geq \left(\frac{\pi^{d/2}|\mathcal{S}|^d}{2^d \Gamma(\frac{d}{2}+1)} \right)^{2/p} \frac{h(d,|\mathcal{S}|,\sigma) + \sqrt{2\tau}}{\sqrt{m}} \right) \leq e^{-\tau}.$$

L^p guarantee

Under the previous assumptions, and $1 \leq p < \infty$:

$$\Lambda^m \left(\|\hat{k} - k\|_{L^p(\mathcal{S})} \geq \left(\frac{\pi^{d/2}|\mathcal{S}|^d}{2^d \Gamma(\frac{d}{2}+1)} \right)^{2/p} \frac{h(d,|\mathcal{S}|,\sigma) + \sqrt{2\tau}}{\sqrt{m}} \right) \leq e^{-\tau}.$$

Hence,

$$\|\hat{k} - k\|_{L^p(\mathcal{S})} = O_{a.s.}\Big(\underbrace{m^{-1/2}|\mathcal{S}|^{2d/p}\sqrt{\log|\mathcal{S}|}}_{L^p(\mathcal{S})\text{-consistency if } \xrightarrow{m\to\infty} 0} \Big).$$

L^p guarantee

Under the previous assumptions, and $1 \le p < \infty$:

$$\Lambda^m \left(\|\hat{k} - k\|_{L^p(\mathcal{S})} \ge \left(\frac{\pi^{d/2}|\mathcal{S}|^d}{2^d \Gamma(\frac{d}{2}+1)} \right)^{2/p} \frac{h(d,|\mathcal{S}|,\sigma) + \sqrt{2\tau}}{\sqrt{m}} \right) \le e^{-\tau}.$$

Hence,

$$\|\hat{k} - k\|_{L^p(\mathcal{S})} = O_{a.s.}\Big(\underbrace{m^{-1/2}|\mathcal{S}|^{2d/p}\sqrt{\log|\mathcal{S}|}}_{L^p(\mathcal{S})\text{-consistency if } \xrightarrow{m\to\infty} 0} \Big).$$

Uniform guarantee: $|\mathcal{S}_m| = e^{m^{\delta<1}}$; now: $\frac{|\mathcal{S}_m|^{2d/p}}{\sqrt{m}} \to 0 \Rightarrow |\mathcal{S}_m| = o(m^{\frac{p}{4d}})$.

Direct L^p guarantee (proof after discussion)

Under the previous assumptions, and $1 < p < \infty$:

$$\Lambda^m\left(\|\hat{k}-k\|_{L^p(\mathcal{S})} \geq \left(\frac{\pi^{d/2}|\mathcal{S}|^d}{2^d\Gamma(\frac{d}{2}+1)}\right)^{2/p}\left(\frac{C_p'}{m^{1-\max\{\frac{1}{2},\frac{1}{p}\}}}+\frac{\sqrt{2\tau}}{\sqrt{m}}\right)\right) \leq e^{-\tau},$$

C_p': universal constant; only p-dependent (not $|\mathcal{S}|$ or m-dep.).

Direct L^p guarantee (proof after discussion)

Under the previous assumptions, and $1 < p < \infty$:

$$\Lambda^m\left(\|\hat{k}-k\|_{L^p(\mathcal{S})} \geq \left(\frac{\pi^{d/2}|\mathcal{S}|^d}{2^d\Gamma(\frac{d}{2}+1)}\right)^{2/p}\left(\frac{C_p'}{m^{1-\max\{\frac{1}{2},\frac{1}{p}\}}}+\frac{\sqrt{2\tau}}{\sqrt{m}}\right)\right) \leq e^{-\tau},$$

C_p': universal constant; only p-dependent (not $|\mathcal{S}|$ or m-dep.).
Note: if $2 \leq p$, then

1. $m^{1-\max\{\frac{1}{2},\frac{1}{p}\}} = \sqrt{m}$ [we got rid of $\sqrt{\log(\mathcal{S})}$],
2. $\|\hat{k}-k\|_{L^p(\mathcal{S}_m)} \stackrel{a.s.}{\to} 0$ if $|\mathcal{S}_m| = o\left(m^{\frac{p}{4d}}\right)$ as $m \to \infty$.

Direct L^p result: High-level idea

1. By the bounded difference property:

$$\|k - \hat{k}\|_{L^p(\mathcal{S})} \le \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \|k - \hat{k}\|_{L^p(\mathcal{S})} + \mathrm{vol}^{2/p}(\mathcal{S}) \sqrt{\frac{2\tau}{m}}.$$

Direct L^p result: High-level idea

1. By the bounded difference property:

$$\|k-\hat{k}\|_{L^p(\mathcal{S})} \le \mathbb{E}_{\boldsymbol{\omega}_{1:m}}\|k-\hat{k}\|_{L^p(\mathcal{S})} + \text{vol}^{2/p}(\mathcal{S})\sqrt{\frac{2\tau}{m}}.$$

2. By $L^p \cong (L^{p'})^*$ $\left(\frac{1}{p}+\frac{1}{p'}=1\right)$, the separability of $L^{p'}(\mathcal{S})$ and symmetrization [van der Vaart and Wellner, 1996, Lemma 2.3.1]:

$$\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\|k-\hat{k}\|_{L^p(\mathcal{S})} \le \frac{2}{m}\mathbb{E}_{\boldsymbol{\omega}_{1:m}} \underbrace{\mathbb{E}_{\boldsymbol{\varepsilon}}\left\|\sum_{i=1}^m \varepsilon_i \cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot\rangle)\right\|_{L^p(\mathcal{S})}}_{=:(*)}.$$

Direct L^p result: High-level idea

1. By the bounded difference property:

$$\|k-\hat{k}\|_{L^p(\mathcal{S})} \le \mathbb{E}_{\boldsymbol{\omega}_{1:m}}\|k-\hat{k}\|_{L^p(\mathcal{S})} + \text{vol}^{2/p}(\mathcal{S})\sqrt{\frac{2\tau}{m}}.$$

2. By $L^p \cong (L^{p'})^*$ $\left(\frac{1}{p}+\frac{1}{p'}=1\right)$, the separability of $L^{p'}(\mathcal{S})$ and symmetrization [van der Vaart and Wellner, 1996, Lemma 2.3.1]:

$$\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\|k-\hat{k}\|_{L^p(\mathcal{S})} \le \frac{2}{m}\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\underbrace{\mathbb{E}_{\boldsymbol{\varepsilon}}\left\|\sum_{i=1}^m \varepsilon_i \cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot\rangle)\right\|_{L^p(\mathcal{S})}}_{=:(*)}.$$

3. Since $L^p(\mathcal{S})$ is of type $\min(2,p)$ $\exists C'_p$ such that

$$(*) \le C'_p\left(\sum_{i=1}^m \|\cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot\rangle)\|_{L^p(\mathcal{S})}^{\min(2,p)}\right)^{\frac{1}{\min(2,p)}}.$$

Direct L^p result: Step-1

$f(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_m) := \|k - \hat{k}\|_{L^p(\mathcal{S})}$ has bounded difference:

$$\hat{k}_i(\mathbf{x}, \mathbf{y}) = \frac{1}{m} \sum_{j \neq i} \cos(\boldsymbol{\omega}_j^T (\mathbf{x} - \mathbf{y})) + \frac{1}{m} \cos(\tilde{\boldsymbol{\omega}}_i^T (\mathbf{x} - \mathbf{y})),$$

$$\sup_{(\boldsymbol{\omega}_i)_{i=1}^m, \tilde{\boldsymbol{\omega}}_i} \left| \|k - \hat{k}\|_{L^p(\mathcal{S})} - \|k - \hat{k}_i\|_{L^p(\mathcal{S})} \right| \leq$$

$$\leq \sup_{(\boldsymbol{\omega}_i)_{i=1}^m, \tilde{\boldsymbol{\omega}}_i} \|\hat{k}_i - \hat{k}\|_{L^p(\mathcal{S})}$$

Direct L^p result: Step-1

$f(\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_m) := \|k-\hat{k}\|_{L^p(\mathcal{S})}$ has bounded difference:

$$\hat{k}_i(\mathbf{x},\mathbf{y}) = \frac{1}{m}\sum_{j\neq i}\cos(\boldsymbol{\omega}_j^T(\mathbf{x}-\mathbf{y})) + \frac{1}{m}\cos(\tilde{\boldsymbol{\omega}}_i^T(\mathbf{x}-\mathbf{y})),$$

$$\sup_{(\boldsymbol{\omega}_i)_{i=1}^m,\tilde{\boldsymbol{\omega}}_i}\left|\|k-\hat{k}\|_{L^p(\mathcal{S})} - \|k-\hat{k}_i\|_{L^p(\mathcal{S})}\right| \leq$$

$$\leq \sup_{(\boldsymbol{\omega}_i)_{i=1}^m,\tilde{\boldsymbol{\omega}}_i}\|\hat{k}_i-\hat{k}\|_{L^p(\mathcal{S})} \leq \frac{2}{m}\sup_{\boldsymbol{\omega}_i}\|\cos(\langle\boldsymbol{\omega}_i,\cdot-\cdot\rangle)\|_{L^p(\mathcal{S})}$$

Direct L^p result: Step-1

$f(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_m) := \|k - \hat{k}\|_{L^p(\mathcal{S})}$ has bounded difference:

$$\hat{k}_i(\mathbf{x}, \mathbf{y}) = \frac{1}{m} \sum_{j \neq i} \cos(\boldsymbol{\omega}_j^T (\mathbf{x} - \mathbf{y})) + \frac{1}{m} \cos(\tilde{\boldsymbol{\omega}}_i^T (\mathbf{x} - \mathbf{y})),$$

$$\sup_{(\boldsymbol{\omega}_i)_{i=1}^m, \tilde{\boldsymbol{\omega}}_i} \left| \|k - \hat{k}\|_{L^p(\mathcal{S})} - \|k - \hat{k}_i\|_{L^p(\mathcal{S})} \right| \leq$$

$$\leq \sup_{(\boldsymbol{\omega}_i)_{i=1}^m, \tilde{\boldsymbol{\omega}}_i} \|\hat{k}_i - \hat{k}\|_{L^p(\mathcal{S})} \leq \frac{2}{m} \sup_{\boldsymbol{\omega}_i} \|\cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot \rangle)\|_{L^p(\mathcal{S})}$$

$$\leq \frac{2}{m} \mathrm{vol}^{2/p}(\mathcal{S}) =: c_r.$$

$\Rightarrow$ We can apply the McDiarmid inequality.

We write $\|\cdot\|_{L^p}$ as a countable sup

Let $1 < p' < \infty$.

- Let $(X, \mathcal{A}, \mu)$, $\mu(X) < \infty$, $\frac{1}{p} + \frac{1}{p'} = 1$. Then

$$\left[L^{p'}(X, \mathcal{A}, \mu)\right]^* = \{F_f : f \in L^p(X, \mathcal{A}, \mu)\},$$

$$F_f(u) = \int_X uf \mathrm{d}\mu,$$

and $\|f\|_{L^p} = \|F_f\| = \sup_{\|g\|_{L^{p'}}=1} |F_f(g)| =: (*)$.

We write $\|\cdot\|_{L^p}$ as a countable sup

Let $1 < p' < \infty$.

- Let $(X, \mathcal{A}, \mu)$, $\mu(X) < \infty$, $\frac{1}{p} + \frac{1}{p'} = 1$. Then

$$\left[L^{p'}(X, \mathcal{A}, \mu)\right]^* = \{F_f : f \in L^p(X, \mathcal{A}, \mu)\},$$

$$F_f(u) = \int_X uf \mathrm{d}\mu,$$

and $\|f\|_{L^p} = \|F_f\| = \sup_{\|g\|_{L^{p'}}=1} |F_f(g)| =: (*)$.

- Moreover, since for $X = \mathcal{S}$, $L^{p'}(\mathcal{S})$ is separable [Cohn, 2013, Prop. 3.4.5] $\Rightarrow \exists\, \mathcal{G} \subseteq S_{L^{p'}(\mathcal{S})}(0, 1)$ countable [Carothers, 2004, Lemma 6.7]: $(*) = \sup_{g \in \mathcal{G}} |F_f(g)|$.

Direct L^p result: Step-2

By the previous rewriting

$$\|k-\hat{k}\|_{L^p(\mathcal{S})} = \left\|F_{k-\hat{k}}\right\| = \sup_{g\in\mathcal{G}}\left|\int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y})\left[k(\mathbf{x},\mathbf{y})-\hat{k}(\mathbf{x},\mathbf{y})\right]\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}\right| =: (*)$$

Direct L^p result: Step-2

By the previous rewriting

$$\|k - \hat{k}\|_{L^p(\mathcal{S})} = \left\|F_{k-\hat{k}}\right\| = \sup_{g\in\mathcal{G}} \left| \int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x}, \mathbf{y}) \left[k(\mathbf{x}, \mathbf{y}) - \hat{k}(\mathbf{x}, \mathbf{y})\right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \right| =: (*)$$

$$\int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x}, \mathbf{y}) \left[k(\mathbf{x}, \mathbf{y}) - \hat{k}(\mathbf{x}, \mathbf{y})\right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$$

Direct L^p result: Step-2

By the previous rewriting

$$\|k-\hat{k}\|_{L^p(\mathcal{S})} = \left\|F_{k-\hat{k}}\right\| = \sup_{g\in\mathcal{G}} \left| \int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y}) \left[k(\mathbf{x},\mathbf{y}) - \hat{k}(\mathbf{x},\mathbf{y})\right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \right| =: (*)$$

$$\int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y}) \left[k(\mathbf{x},\mathbf{y}) - \hat{k}(\mathbf{x},\mathbf{y})\right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$$
$$= \int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y}) \left[\int_{\mathbb{R}^d} \cos(\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y}))\mathrm{d}(\Lambda-\Lambda_m)(\boldsymbol{\omega})\right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$$

Direct L^p result: Step-2

By the previous rewriting

$$\|k-\hat{k}\|_{L^p(\mathcal{S})} = \left\|F_{k-\hat{k}}\right\| = \sup_{g\in\mathcal{G}}\left|\int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y})\left[k(\mathbf{x},\mathbf{y})-\hat{k}(\mathbf{x},\mathbf{y})\right]\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}\right| =: (*)$$

$$\int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y})\left[k(\mathbf{x},\mathbf{y})-\hat{k}(\mathbf{x},\mathbf{y})\right]\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$$

$$= \int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y})\left[\int_{\mathbb{R}^d}\cos(\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y}))\mathrm{d}(\Lambda-\Lambda_m)(\boldsymbol{\omega})\right]\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$$

$$= \int_{\mathbb{R}^d}\underbrace{\int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y})\cos(\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y}))\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}}_{=:\tilde{g}_g(\boldsymbol{\omega}):\ \text{measurable}\ \Leftarrow\ \text{dominated convergence}}\ \mathrm{d}(\Lambda-\Lambda_m)(\boldsymbol{\omega}) \Rightarrow$$

Direct L^p result: Step-2

By the previous rewriting

$$\|k-\hat{k}\|_{L^p(\mathcal{S})} = \left\|F_{k-\hat{k}}\right\| = \sup_{g\in\mathcal{G}} \left| \int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y}) \left[k(\mathbf{x},\mathbf{y}) - \hat{k}(\mathbf{x},\mathbf{y})\right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \right| =: (*)$$

$$\begin{aligned}
&\int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y}) \left[k(\mathbf{x},\mathbf{y}) - \hat{k}(\mathbf{x},\mathbf{y})\right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&= \int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y}) \left[\int_{\mathbb{R}^d} \cos(\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y}))\mathrm{d}(\Lambda-\Lambda_m)(\boldsymbol{\omega})\right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\
&= \int_{\mathbb{R}^d} \underbrace{\int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y}) \cos(\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y}))\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}}_{=:\tilde{g}_g(\boldsymbol{\omega}):\ \text{measurable} \Leftarrow \text{dominated convergence}} \mathrm{d}(\Lambda-\Lambda_m)(\boldsymbol{\omega}) \Rightarrow
\end{aligned}$$

$$(*) = \sup_{\tilde{g}\in\tilde{\mathcal{G}}:=\{\tilde{g}_g:g\in\mathcal{G}\}} \left|(\Lambda-\Lambda_m)\tilde{g}\right|,$$

Direct L^p result: Step-2

By symmetrization [(a)] we get

$$\mathbb{E}_{\boldsymbol{\omega}_{1:m}} \|k - \hat{k}\|_{L^p(\mathcal{S})} \overset{(a)}{\leq} 2\mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{\tilde{g} \in \tilde{\mathcal{G}}} \left| \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i \tilde{g}(\boldsymbol{\omega}_i) \right|$$

Direct L^p result: Step-2

By symmetrization [(a)], $\tilde{g}$ def. [(b)] we get

$$\mathbb{E}_{\boldsymbol{\omega}_{1:m}} \|k - \hat{k}\|_{L^p(\mathcal{S})} \overset{(a)}{\leq} 2\mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{\tilde{g} \in \tilde{\mathcal{G}}} \left| \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i \tilde{g}(\boldsymbol{\omega}_i) \right|$$

$$\overset{(b)}{=} \frac{2}{m} \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{g \in \mathcal{G}} \left| \sum_{i=1}^{m} \varepsilon_i \int_{\mathcal{S} \times \mathcal{S}} g(\mathbf{x}, \mathbf{y}) \cos\left(\boldsymbol{\omega}_i^T (\mathbf{x} - \mathbf{y})\right) \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{y} \right|$$

Direct L^p result: Step-2

By symmetrization [(a)], $\tilde{g}$ def. [(b)] we get

$$\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\|k-\hat{k}\|_{L^p(\mathcal{S})} \overset{(a)}{\leq} 2\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{\tilde{g}\in\tilde{\mathcal{G}}} \left| \frac{1}{m}\sum_{i=1}^m \varepsilon_i \tilde{g}(\boldsymbol{\omega}_i) \right|$$

$$\overset{(b)}{=} \frac{2}{m}\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{g\in\mathcal{G}} \left| \sum_{i=1}^m \varepsilon_i \int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y}) \cos\left(\boldsymbol{\omega}_i^T(\mathbf{x}-\mathbf{y})\right) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \right|$$

$$= \frac{2}{m}\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{g\in\mathcal{G}} \left| \int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y}) \left[\sum_{i=1}^m \varepsilon_i \cos\left(\boldsymbol{\omega}_i^T(\mathbf{x}-\mathbf{y})\right)\right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \right|$$

Direct L^p result: Step-2

By symmetrization [(a)], $\tilde{g}$ def. [(b)] and $L^p \cong (L^{p'})^*$ [(c)], we get

$$\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\|k-\hat{k}\|_{L^p(\mathcal{S})} \overset{(a)}{\leq} 2\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{\tilde{g}\in\tilde{\mathcal{G}}}\left|\frac{1}{m}\sum_{i=1}^m \varepsilon_i \tilde{g}(\boldsymbol{\omega}_i)\right|$$

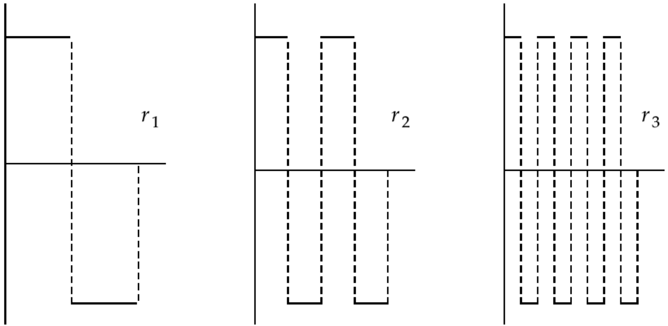
$$\overset{(b)}{=} \frac{2}{m}\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{g\in\mathcal{G}}\left|\sum_{i=1}^m \varepsilon_i \int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y}) \cos\left(\boldsymbol{\omega}_i^T(\mathbf{x}-\mathbf{y})\right) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}\right|$$

$$= \frac{2}{m}\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{g\in\mathcal{G}}\left|\int_{\mathcal{S}\times\mathcal{S}} g(\mathbf{x},\mathbf{y}) \left[\sum_{i=1}^m \varepsilon_i \cos\left(\boldsymbol{\omega}_i^T(\mathbf{x}-\mathbf{y})\right)\right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}\right|$$

$$\overset{(c)}{=} \frac{2}{m}\mathbb{E}_{\boldsymbol{\omega}_{1:m}}\mathbb{E}_{\boldsymbol{\varepsilon}}\left\|\sum_{i=1}^m \varepsilon_i \cos(\langle\boldsymbol{\omega}_i, \cdot - \cdot\rangle)\right\|_{L^p(\mathcal{S})}.$$

Direct L^p result: Step-3

Rademacher functions: $r_j(s) = sgn\left(\sin\left(2^j \pi s\right)\right) \in L^2[0,1]$ $(j = 1, \ldots)$.



Direct L^p result: Step-3

Properties of Rademacher functions:

1. ONS in $L^2[0,1]$.

Direct L^p result: Step-3

Properties of Rademacher functions:

1. ONS in $L^2[0,1]$.
2. $[r_1(t);\dots;r_m(t)] = [\epsilon_1;\dots;\epsilon_m] \in \{-1,1\}^m$ Rademacher vector, where $t \sim U[0,1] \Rightarrow$

$$\mathbb{E}_{\boldsymbol{\varepsilon}} \left\| \sum_{j=1}^{m} \varepsilon_j f_j \right\| = \int_0^1 \left\| \sum_{j=1}^{m} r_j(s) f_j \right\| \mathrm{d}s.$$

Direct L^p result: Step-3

A $(Z, \|\cdot\|)$ Banach space is of type $q \in (1, 2]$ if $\exists C \in \mathbb{R}$

$$\int_0^1 \left\| \sum_{j=1}^m r_j(s) f_j \right\| \mathrm{d}s \le C \left(\sum_{j=1}^m \|f_j\|^q \right)^{\frac{1}{q}}, \forall m, \forall \{f_j\}_{j=1}^m \subseteq Z.$$

Direct L^p result: Step-3

A $(Z, \|\cdot\|)$ Banach space is of type $q \in (1, 2]$ if $\exists C \in \mathbb{R}$

$$\int_0^1 \left\| \sum_{j=1}^m r_j(s) f_j \right\| \mathrm{d}s \leq C \left(\sum_{j=1}^m \|f_j\|^q \right)^{\frac{1}{q}}, \forall m, \forall \{f_j\}_{j=1}^m \subseteq Z.$$

Notes:

1. q choice: $\forall$ ($\nexists$) B-space is of type 1 (> 2).

Direct L^p result: Step-3

A $(Z, \|\cdot\|)$ Banach space is of type $q \in (1, 2]$ if $\exists C \in \mathbb{R}$

$$\int_0^1 \left\| \sum_{j=1}^m r_j(s) f_j \right\| \mathrm{d}s \leq C \left(\sum_{j=1}^m \|f_j\|^q \right)^{\frac{1}{q}}, \forall m, \forall \{f_j\}_{j=1}^m \subseteq Z.$$

Notes:

1. q choice: $\forall$ ($\nexists$) B-space is of type 1 (> 2).
2. $\forall$ Hilbert space is of type 2.

Direct L^p result: Step-3

A $(Z, \|\cdot\|)$ Banach space is of type $q \in (1, 2]$ if $\exists C \in \mathbb{R}$

$$\int_0^1 \left\| \sum_{j=1}^m r_j(s) f_j \right\| \mathrm{d}s \leq C \left(\sum_{j=1}^m \|f_j\|^q \right)^{\frac{1}{q}}, \forall m, \forall \{f_j\}_{j=1}^m \subseteq Z.$$

Notes:

1. q choice: $\forall$ ($\nexists$) B-space is of type 1 (> 2).
2. $\forall$ Hilbert space is of type 2.
3. $Z = L^p(X, \mathcal{A}, \mu)$ is of type $q = \min(2, p)$ [Lindenstrauss and Tzafriri, 1979, page 73] $\Rightarrow$

Direct L^p result: Step-3

$\exists C'_p$ such that

$$\mathbb{E}_{\varepsilon}\left\|\sum_{i=1}^{m}\varepsilon_i \cos(\langle\omega_i,\cdot-\cdot\rangle)\right\|_{L^p(\mathcal{S})} \le C'_p\left(\sum_{i=1}^{m}\|\cos(\langle\omega_i,\cdot-\cdot\rangle)\|_{L^p(\mathcal{S})}^{\min(2,p)}\right)^{\frac{1}{\min(2,p)}} =: (*)$$

$$\sum_{i=1}^{m}\|\cos(\langle\omega_i,\cdot-\cdot\rangle)\|_{L^p(\mathcal{S})}^{\min(2,p)} =$$

Direct L^p result: Step-3

$\exists C'_p$ such that

$$\mathbb{E}_{\varepsilon}\left\|\sum_{i=1}^{m}\varepsilon_i \cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot\rangle)\right\|_{L^p(\mathcal{S})} \le C'_p \left(\sum_{i=1}^{m}\|\cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot\rangle)\|_{L^p(\mathcal{S})}^{\min(2,p)}\right)^{\frac{1}{\min(2,p)}} =: (*)$$

$$\sum_{i=1}^{m}\|\cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot\rangle)\|_{L^p(\mathcal{S})}^{\min(2,p)} = \sum_{i=1}^{m}\left(\int_{\mathcal{S}\times\mathcal{S}}\underbrace{\left|\cos(\boldsymbol{\omega}_i^T(\mathbf{x}-\mathbf{y}))\right|^p}_{\le 1} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}\right)^{\frac{\min(2,p)}{p}}$$

Direct L^p result: Step-3

$\exists C'_p$ such that

$$\mathbb{E}_{\boldsymbol{\varepsilon}} \left\| \sum_{i=1}^m \varepsilon_i \cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot \rangle) \right\|_{L^p(\mathcal{S})} \le C'_p \left(\sum_{i=1}^m \| \cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot \rangle) \|_{L^p(\mathcal{S})}^{\min(2,p)} \right)^{\frac{1}{\min(2,p)}} =: (*)$$

$$\sum_{i=1}^m \| \cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot \rangle) \|_{L^p(\mathcal{S})}^{\min(2,p)} = \sum_{i=1}^m \left(\int_{\mathcal{S}\times\mathcal{S}} \underbrace{\left| \cos(\boldsymbol{\omega}_i^T(\mathbf{x}-\mathbf{y})) \right|^p}_{\le 1} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \right)^{\frac{\min(2,p)}{p}}$$

$$\le m \left[\mathrm{vol}^2(\mathcal{S}) \right]^{\frac{\min(2,p)}{p}} \Rightarrow$$

Direct L^p result: Step-3

$\exists C'_p$ such that

$$\mathbb{E}_{\varepsilon}\left\|\sum_{i=1}^m \varepsilon_i \cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot \rangle)\right\|_{L^p(\mathcal{S})} \le C'_p \left(\sum_{i=1}^m \|\cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot \rangle)\|_{L^p(\mathcal{S})}^{\min(2,p)}\right)^{\frac{1}{\min(2,p)}} =: (*)$$

$$\sum_{i=1}^m \|\cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot \rangle)\|_{L^p(\mathcal{S})}^{\min(2,p)} = \sum_{i=1}^m \left(\int_{\mathcal{S}\times\mathcal{S}} \underbrace{\left|\cos(\boldsymbol{\omega}_i^T(\mathbf{x}-\mathbf{y}))\right|^p}_{\le 1} \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}\right)^{\frac{\min(2,p)}{p}}$$

$$\le m \left[\mathrm{vol}^2(\mathcal{S})\right]^{\frac{\min(2,p)}{p}} \Rightarrow$$

$$(*) \le C'_p m^{\frac{1}{\min(2,p)} = \max\{\frac{1}{2},\frac{1}{p}\}} \mathrm{vol}^{2/p}(\mathcal{S}).$$

Summary

Finite sample

- $L^\infty(\mathcal{S})$ guarantee $\xrightarrow{\text{spec.}} |\mathcal{S}_m| = e^{o(m)}$ – optimal!
- $L^p(\mathcal{S})$ results ($\Leftarrow$ uniform, type of L^p).

Summary

Finite sample

- $L^\infty(\mathcal{S})$ guarantee $\xrightarrow{\text{spec.}} |\mathcal{S}_m| = e^{o(m)}$ – optimal!
- $L^p(\mathcal{S})$ results ($\Leftarrow$ uniform, type of L^p).

Thank you for the attention!

Contents

- Borel-Cantelli lemma.
- McDiarmid inequality.
- $L^\infty(\mathcal{S})$ is *not* separable.

Borel-Cantelli lemma

- Assume: $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$.
- Then $\mathbb{P}(\infty$-ly many of them occur$) = 0$. Formally,

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0,$$

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

McDiarmid inequality [Shawe-Taylor and Cristianini, 2004]

Let $\omega_1, \ldots, \omega_m \in D$ be independent r.v.-s, and $f : D^m \to \mathbb{R}$ satisfy the bounded diff. property $(\forall r)$:

$$\sup_{u_1,\ldots,u_m,u'_r \in D} \left| f(u_1, \ldots, u_m) - f(u_1, \ldots, u_{r-1}, u'_r, u_{r+1}, \ldots, u_m) \right| \leq c_r.$$

Then for $\forall \beta > 0$

$$\mathbb{P}\left(f(\omega_1, \ldots, \omega_m) - \mathbb{E}\left[f(\omega_1, \ldots, \omega_m)\right] \geq \beta\right) \leq e^{-\frac{2\beta^2}{\sum_{r=1}^m c_r^2}}.$$

Note: specifically, if $c = c_r$ $(\forall r)$, $\tau = \frac{2\epsilon^2}{\sum_{r=1}^m c_r^2} = \frac{2\epsilon^2}{mc^2} \Leftrightarrow \epsilon = c\sqrt{\frac{\tau m}{2}}$ gives $\mathbb{P}\left(f(X_1, \ldots, X_m) < \mathbb{E}\left[f(X_1, \ldots, X_m)\right] + c\sqrt{\frac{\tau m}{2}}\right) \geq 1 - e^{-\tau}$.

$L^\infty(\mathcal{S})$ is *not* separable

- Assume that $0 \in \mathcal{S}$.
- Take $S := \{I_{B(0,r)}\}_{r>0} \subseteq L^\infty(\mathcal{S})$.
- $|S| >$ countable, and for $\forall s_1 \neq s_2 \in S$: $\|s_1 - s_2\|_{L^\infty(\mathcal{S})} = 1$.

Bousquet, O. (2003).
New approaches to statistical learning theory.
Annals of the Institute of Statistical Mathematics, 55:371–389.

Carothers, N. L. (2004).
A Short Course on Banach Space Theory.
Cambridge University Press.

Cohn, D. L. (2013).
Measure Theory: Second Edition.
Birkhäuser Basel.

Csörgő, S. and Totik, V. (1983).
On how long interval is the empirical characteristic function uniformly consistent?
Acta Sci. Math. (Szeged), 45:141–149.

Lindenstrauss, J. and Tzafriri, L. (1979).
Classical Banach Spaces II – Function Spaces.
Springer-Verlag.

Rahimi, A. and Recht, B. (2007).
Random features for large-scale kernel machines.
In *Neural Information Processing Systems (NIPS)*, pages 1177–1184.

Shawe-Taylor, J. and Cristianini, N. (2004).
Kernel Methods for Pattern Analysis.
Cambridge University Press.

Steinwart, I. and Christmann, A. (2008).
Support Vector Machines.
Springer.

van de Geer, S. A. (2009).
Empirical Processes in M-Estimation.
Cambridge University Press.

van der Vaart, A. W. and Wellner, J. A. (1996).
Weak Convergence and Empirical Processes.
Springer-Verlag.