Optimal Uniform and L^p Rates for Random Fourier Features

Zoltán Szabó

Joint work with Bharath K. Sriperumbudur (PSU)

Gatsby Unit, Research Talk
September 7, 2015

Recap

- Given:

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

- $s^{\mathbf{p},\mathbf{q}}(\mathbf{x},\mathbf{y})$: Monte-Carlo estimator of $\partial^{\mathbf{p},\mathbf{q}} k(\mathbf{x},\mathbf{y})$ using $(\boldsymbol{\omega}_j)_{j=1}^m \stackrel{i.i.d.}{\sim} \Lambda$.

Recap

- Given:

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

- $s^{\mathbf{p},\mathbf{q}}(\mathbf{x},\mathbf{y})$: Monte-Carlo estimator of $\partial^{\mathbf{p},\mathbf{q}}k(\mathbf{x},\mathbf{y})$ using $(\boldsymbol{\omega}_j)_{j=1}^m \overset{i.i.d.}{\sim} \Lambda$.
- **Last time**:

$$\left\|\partial^{\mathbf{p},\mathbf{q}}k - s^{\mathbf{p},\mathbf{q}}\right\|_{L^\infty(\mathcal{K})} = \mathcal{O}_{a.s.}\left(\frac{\sqrt{|\mathcal{K}|}}{\sqrt{m}}\right).$$

Derivatives: '*supp*(Λ) is bounded' requirement.

Today: one-page summary

1. Tighter L^∞ guarantee in terms of $|\mathcal{K}|$:

$$\left\|\partial^{\mathbf{p},\mathbf{q}} k - s^{\mathbf{p},\mathbf{q}}\right\|_{L^\infty(\mathcal{K})} = \mathcal{O}_{a.s.}\left(\frac{\sqrt{\log|\mathcal{K}|}}{\sqrt{m}}\right)$$

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

Today: one-page summary

1. Tighter L^∞ guarantee in terms of $|\mathcal{K}|$:

$$\left\|\partial^{\mathbf{p},\mathbf{q}} k - s^{\mathbf{p},\mathbf{q}}\right\|_{L^\infty(\mathcal{K})} = \mathcal{O}_{a.s.}\left(\frac{\sqrt{\log|\mathcal{K}|}}{\sqrt{m}}\right)$$

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

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

Today: one-page summary

1. Tighter L^{∞} guarantee in terms of $|\mathcal{K}|$:

$$\left\|\partial^{\mathbf{p},\mathbf{q}}k - s^{\mathbf{p},\mathbf{q}}\right\|_{L^{\infty}(\mathcal{K})} = \mathcal{O}_{a.s.}\left(\frac{\sqrt{\log|\mathcal{K}|}}{\sqrt{m}}\right)$$

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

2. Finite sample L^r guarantees, $r \in [1, \infty)$.
3. Moment constraints on Λ are enough (example: RBF k).

Dissemination

- **Theoretical foundations**: Bharath K. Sriperumbudur, Zoltán Szabó (contributed equally). Optimal Rates for Random Fourier Features. In NIPS-2015, accepted [for spotlight presentation - 3.65%].

Dissemination

- **Theoretical foundations**: Bharath K. Sriperumbudur, Zoltán Szabó (contributed equally). Optimal Rates for Random Fourier Features. In NIPS-2015, accepted [for spotlight presentation - 3.65%].
- **Infinite dimensional exponential family fitting application**: Heiko Strathmann, Dino Sejdinovic, Samuel Livingston, Zoltán Szabó, Arthur Gretton. Gradient-free Hamiltonian Monte Carlo with Efficient Kernel Exponential Families. In NIPS-2015, accepted.

L^∞ guarantee

[Csörgő and Totik, 1983]'s asymptotic result:

1. $|\mathcal{K}_m| = e^{o(m)}$ is the optimal rate for a.s. convergence,
2. For faster growing $|\mathcal{K}_m|$: even convergence is probability fails.

L^∞ guarantee

[Csörgő and Totik, 1983]'s asymptotic result:

1. $|\mathcal{K}_m| = e^{o(m)}$ is the optimal rate for a.s. convergence,
2. For faster growing $|\mathcal{K}_m|$: even convergence is probability fails.

Goal:

1. finite sample L^∞ guarantee,
2. which implies this optimal rate.

L^∞ guarantee

We saw $[h_a = cos^{(a)}]$:

$$\left\|\partial^{\mathbf{p},\mathbf{q}} k - s^{\mathbf{p},\mathbf{q}}\right\|_{L^\infty(\mathcal{K})} \precsim \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) + \frac{1}{\sqrt{m}},$$

L^∞ guarantee

We saw $[h_a = cos^{(a)}]$:

$$\left\|\partial^{\mathbf{p},\mathbf{q}} k - s^{\mathbf{p},\mathbf{q}}\right\|_{L^\infty(\mathcal{K})} \lesssim \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) + \frac{1}{\sqrt{m}},$$

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

L^∞ guarantee

We saw $[h_a = cos^{(a)}]$:

$$\left\|\partial^{\mathbf{p},\mathbf{q}} k - s^{\mathbf{p},\mathbf{q}}\right\|_{L^\infty(\mathcal{K})} \lesssim \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) + \frac{1}{\sqrt{m}},$$

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

$$\lesssim \frac{1}{\sqrt{m}} \int_0^{|\mathcal{G}|_{L^2(\Lambda_m)}} \sqrt{\log \mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)} \mathrm{d}r,$$

L^∞ guarantee

We saw $[h_a = cos^{(a)}]$:

$$\left\|\partial^{\mathbf{p},\mathbf{q}} k - s^{\mathbf{p},\mathbf{q}}\right\|_{L^\infty(\mathcal{K})} \lesssim \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) + \frac{1}{\sqrt{m}},$$

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

$$\lesssim \frac{1}{\sqrt{m}} \int_0^{|\mathcal{G}|_{L^2(\Lambda_m)}} \sqrt{\log \mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)} \mathrm{d}r,$$

$$\mathcal{G} = \{g_{\mathbf{z}}(\boldsymbol{\omega}) = \boldsymbol{\omega}^{\mathbf{p}}(-\boldsymbol{\omega})^{\mathbf{q}} h_{|\mathbf{p}+\mathbf{q}|}(\boldsymbol{\omega}^T \mathbf{z}) : \mathbf{z} \in \mathcal{K}_\Delta\},$$

L^∞ guarantee

We saw $[h_a = cos^{(a)}]$:

$$\left\|\partial^{\mathbf{p},\mathbf{q}} k - s^{\mathbf{p},\mathbf{q}}\right\|_{L^\infty(\mathcal{K})} \lesssim \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) + \frac{1}{\sqrt{m}},$$

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

$$\lesssim \frac{1}{\sqrt{m}} \int_0^{|\mathcal{G}|_{L^2(\Lambda_m)}} \sqrt{\log \mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)} \mathrm{d}r,$$

$$\mathcal{G} = \left\{g_{\mathbf{z}}(\omega) = \omega^{\mathbf{p}}(-\omega)^{\mathbf{q}} h_{|\mathbf{p}+\mathbf{q}|}(\omega^T \mathbf{z}) : \mathbf{z} \in \mathcal{K}_\Delta\right\},$$

$$\mathcal{N}\left(\mathcal{G}, L^2(\Lambda_m), r\right) \leq \left(\frac{4|\mathcal{K}| A_{\mathbf{p},\mathbf{q}}}{r} + 1\right)^d,$$

L^∞ guarantee

We saw $[h_a = cos^{(a)}]$:

$$\|\partial^{\mathbf{p},\mathbf{q}} k - s^{\mathbf{p},\mathbf{q}}\|_{L^\infty(\mathcal{K})} \lesssim \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) + \frac{1}{\sqrt{m}},$$

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

$$\lesssim \frac{1}{\sqrt{m}} \int_0^{|\mathcal{G}|_{L^2(\Lambda_m)}} \sqrt{\log \mathcal{N}(\mathcal{G}, L^2(\Lambda_m), r)} \mathrm{d}r,$$

$$\mathcal{G} = \{g_{\mathbf{z}}(\omega) = \omega^{\mathbf{p}}(-\omega)^{\mathbf{q}} h_{|\mathbf{p}+\mathbf{q}|}(\omega^T \mathbf{z}) : \mathbf{z} \in \mathcal{K}_\Delta\},$$

$$\mathcal{N}\left(\mathcal{G}, L^2(\Lambda_m), r\right) \leq \left(\frac{4|\mathcal{K}|A_{\mathbf{p},\mathbf{q}}}{r} + 1\right)^d,$$

$$A_{\mathbf{p},\mathbf{q}} = \sqrt{\frac{1}{m} \sum_{j=1}^m \left|\omega_j^{2(\mathbf{p}+\mathbf{q})}\right| \|\omega_j\|_2^2}.$$

L^∞ guarantee

Key observation:

$$\log\left[\mathcal{N}\left(\mathcal{G}, L^2(\Lambda_m), r\right)\right] \le d \log\left(\frac{4|\mathcal{K}|\sqrt{\frac{1}{m}\sum_{j=1}^m \left|\boldsymbol{\omega}_j^{2(\mathbf{p}+\mathbf{q})}\right| \|\boldsymbol{\omega}_j\|_2^2}}{r} + 1\right),$$

$\log(u+1) \le u$ was applied $\Rightarrow |\mathcal{K}|$.

L^∞ guarantee: $T_{\mathbf{p},\mathbf{q}} = \sup_{\omega \in supp(\Lambda)} |\boldsymbol{\omega}^{\mathbf{p}+\mathbf{q}}|$

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

L^∞ guarantee: $T_{\mathbf{p},\mathbf{q}} = \sup_{\omega \in supp(\Lambda)} \left|\omega^{\mathbf{p}+\mathbf{q}}\right|$

$$\mathcal{R}\left(\mathcal{G}, \omega_{1:m}\right) \le \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^{2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}} \sqrt{\log\left(\frac{4|\mathcal{K}|A_{\mathbf{p},\mathbf{q}}}{r} + 1\right)} \mathrm{d}r$$

$$\overset{(a)}{\le} \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^{2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}} \sqrt{\log\left(\frac{4|\mathcal{K}|A_{\mathbf{p},\mathbf{q}} + 2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}}{r}\right)} \mathrm{d}r$$

(a): $r \le 2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}$

L^∞ guarantee: $T_{\mathbf{p},\mathbf{q}} = \sup_{\omega \in supp(\Lambda)} \left|\boldsymbol{\omega}^{\mathbf{p}+\mathbf{q}}\right|$

$$\mathcal{R}\left(\mathcal{G}, \boldsymbol{\omega}_{1:m}\right) \leq \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^{2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}} \sqrt{\log\left(\frac{4|\mathcal{K}|A_{\mathbf{p},\mathbf{q}}}{r} + 1\right)} \mathrm{d}r$$

$$\overset{(a)}{\leq} \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^{2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}} \sqrt{\log\left(\frac{4|\mathcal{K}|A_{\mathbf{p},\mathbf{q}} + 2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}}{r}\right)} \mathrm{d}r$$

(a): $r \leq 2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}$, (b): $2|\mathcal{K}|A_{\mathbf{p},\mathbf{q}} + \sqrt{T_{2\mathbf{p},2\mathbf{q}}} \leq \left(2|\mathcal{K}| + \sqrt{T_{2\mathbf{p},2\mathbf{q}}}\right)\left(A_{\mathbf{p},\mathbf{q}} + 1\right)$.

L^∞ guarantee: $T_{\mathbf{p},\mathbf{q}} = \sup_{\omega \in supp(\Lambda)} \left|\omega^{\mathbf{p}+\mathbf{q}}\right|$

$$\mathcal{R}\left(\mathcal{G}, \omega_{1:m}\right) \le \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^{2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}} \sqrt{\log\left(\frac{4|\mathcal{K}|A_{\mathbf{p},\mathbf{q}}}{r} + 1\right)} \mathrm{d}r$$

$$\overset{(a)}{\le} \frac{8\sqrt{2d}}{\sqrt{m}} \int_0^{2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}} \sqrt{\log\left(\frac{4|\mathcal{K}|A_{\mathbf{p},\mathbf{q}} + 2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}}{r}\right)} \mathrm{d}r$$

$$\overset{(b)}{\le} \frac{8\sqrt{2d}}{\sqrt{m}} \left(\int_0^{2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}} \sqrt{\log \frac{2\left(2|\mathcal{K}| + \sqrt{T_{2\mathbf{p},2\mathbf{q}}}\right)}{r}} \mathrm{d}r + 2\sqrt{T_{2\mathbf{p},2\mathbf{q}} \log(A_{\mathbf{p},\mathbf{q}} + 1)} \right).$$

(a): $r \le 2\sqrt{T_{2\mathbf{p},2\mathbf{q}}}$, (b): $2|\mathcal{K}|A_{\mathbf{p},\mathbf{q}} + \sqrt{T_{2\mathbf{p},2\mathbf{q}}} \le \left(2|\mathcal{K}| + \sqrt{T_{2\mathbf{p},2\mathbf{q}}}\right)(A_{\mathbf{p},\mathbf{q}} + 1)$.

L^∞ guarantee

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

(a): change of variables, $B_{\mathbf{p},\mathbf{q}} := \frac{2|\mathcal{K}|}{\sqrt{T_{2\mathbf{p},2\mathbf{q}}}}$

L^∞ guarantee

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

(a): change of variables, $B_{\mathbf{p},\mathbf{q}} := \frac{2|\mathcal{X}|}{\sqrt{T_{2\mathbf{p},2\mathbf{q}}}}$, (b):
$\int_0^1 \sqrt{\log\frac{a}{\epsilon}}\,\mathrm{d}\epsilon \leq \sqrt{\log a} + \frac{1}{2\sqrt{\log a}}$ $(a > 1)$.

L^∞ guarantee

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

$$\overset{(b)}{\leq} \frac{16\sqrt{2d}}{\sqrt{m}}\sqrt{T_{2\mathbf{p},2\mathbf{q}}}\left[\sqrt{\log(B_{\mathbf{p},\mathbf{q}}+1)} + \frac{1}{2\sqrt{\log(B_{\mathbf{p},\mathbf{q}}+1)}} + \right.$$
$$\left. + \sqrt{\log\left(A_{\mathbf{p},\mathbf{q}}+1\right)}\right],$$

(a): change of variables, $B_{\mathbf{p},\mathbf{q}} := \frac{2|\mathcal{K}|}{\sqrt{T_{2\mathbf{p},2\mathbf{q}}}}$, (b): $\int_0^1 \sqrt{\log\frac{a}{\epsilon}}\,\mathrm{d}\epsilon \leq \sqrt{\log a} + \frac{1}{2\sqrt{\log a}}$ $(a > 1)$. $\Rightarrow$

L^∞ result for $\mathbf{p} = \mathbf{q} = \mathbf{0}$ (k)

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

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

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

Consequence-1 (Borel-Cantelli lemma)

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

Consequence-1 (Borel-Cantelli lemma)

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

Consequence-1 (Borel-Cantelli lemma)

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

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

Idea:

- Note that

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

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

Idea:

- Note that

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

- $\mathrm{vol}(\mathcal{K}) \le \mathrm{vol}(B)$, where $B := \left\{ \mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 \le \frac{|\mathcal{K}|}{2} \right\}$,

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

Idea:

- Note that

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

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

L^r guarantee

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

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

L^r guarantee

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

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

Hence,

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

L^r guarantee

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

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

Hence,

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

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

Direct L^r guarantee (proof after discussion)

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

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

C'_r: universal constant; only r-dependent (not $|\mathcal{K}|$ or m-dep.).

Direct L^r guarantee (proof after discussion)

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

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

C'_r: universal constant; only r-dependent (not $|\mathcal{K}|$ or m-dep.).
Note: if $2 \leq r$, then

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

Direct L^r result: High-level idea

1. By the bounded difference property:

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

Direct L^r result: High-level idea

1. By the bounded difference property:

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

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

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

Direct L^r result: High-level idea

1. By the bounded difference property:

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

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

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

3. Since $L^r(\mathcal{X})$ is of type $\min(2,r)$ $\exists C_r'$ such that

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

Direct L^r result: Step-1

$f(\omega_1, \dots, \omega_m) := \|k - \hat{k}\|_{L^r(\mathcal{X})}$ has bounded difference:

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

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

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

Direct L^r result: Step-1

$f(\omega_1,\ldots,\omega_m) := \|k - \hat{k}\|_{L^r(\mathcal{X})}$ has bounded difference:

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

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

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

Direct L^r result: Step-1

$f(\omega_1,\ldots,\omega_m) := \|k-\hat{k}\|_{L^r(\mathcal{K})}$ has bounded difference:

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

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

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

$$\leq \frac{2}{m}\mathrm{vol}^{2/r}(\mathcal{K}) =: c_m.$$

$\Rightarrow$ We can apply the McDiarmid inequality.

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

Let $1 < r' < \infty$.

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

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

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

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

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

Let $1 < r' < \infty$.

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

$$\left[L^{r'}(X, \mathcal{A}, \mu)\right]^* = \{F_f : f \in L^r(X, \mathcal{A}, \mu)\},$$
$$F_f(u) = \int_X uf \, \mathrm{d}\mu,$$

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

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

Direct L^r result: Step-2

By the previous rewriting

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

Direct L^r result: Step-2

By the previous rewriting

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

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

Direct L^r result: Step-2

By the previous rewriting

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

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

Direct L^r result: Step-2

By the previous rewriting

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

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

Direct L^r result: Step-2

By the previous rewriting

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

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

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

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

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

Direct L^r result: Step-2

By symmetrization [(a)] we get

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

Direct L^r result: Step-2

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

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

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

Direct L^r result: Step-2

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

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

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

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

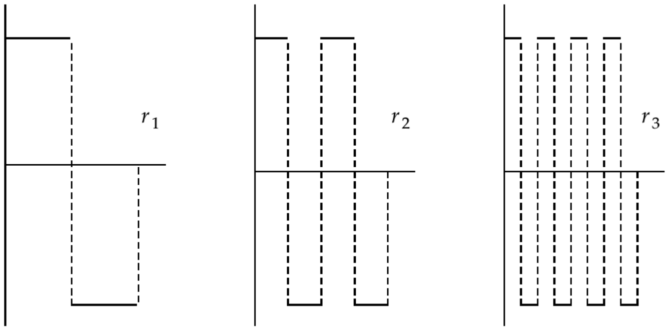
Direct L^r result: Step-2

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

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\omega}_{1:m}} \|k - \hat{k}\|_{L^r(\mathcal{K})} &\overset{(a)}{\leq} 2\mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{\tilde{g}\in\tilde{\mathcal{G}}} \left| \frac{1}{m} \sum_{i=1}^m \varepsilon_i \tilde{g}(\boldsymbol{\omega}_i) \right| \\
&\overset{(b)}{=} \frac{2}{m} \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{g\in\mathcal{G}} \left| \sum_{i=1}^m \varepsilon_i \int_{\mathcal{K}\times\mathcal{K}} g(\mathbf{x},\mathbf{y}) \cos\left(\boldsymbol{\omega}_i^T(\mathbf{x}-\mathbf{y})\right) \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \right| \\
&= \frac{2}{m} \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathbb{E}_{\boldsymbol{\varepsilon}} \sup_{g\in\mathcal{G}} \left| \int_{\mathcal{K}\times\mathcal{K}} g(\mathbf{x},\mathbf{y}) \left[\sum_{i=1}^m \varepsilon_i \cos\left(\boldsymbol{\omega}_i^T(\mathbf{x}-\mathbf{y})\right) \right] \mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \right| \\
&\overset{(c)}{=} \frac{2}{m} \mathbb{E}_{\boldsymbol{\omega}_{1:m}} \mathbb{E}_{\boldsymbol{\varepsilon}} \left\| \sum_{i=1}^m \varepsilon_i \cos(\langle \boldsymbol{\omega}_i, \cdot - \cdot \rangle) \right\|_{L^r(\mathcal{K})}.
\end{aligned}$$

Direct L^r 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^r result: Step-3

Properties of Rademacher functions:

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

Direct L^r result: Step-3

Properties of Rademacher functions:

1. ONS in $L^2[0,1]$.
2. $[r_1(t); \ldots; r_m(t)] = [\epsilon_1; \ldots; \epsilon_m] \in \{-1,1\}^m$ 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^r 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^r 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^r result: Step-3

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

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

Notes:

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

Direct L^r result: Step-3

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

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

Notes:

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

Direct L^r result: Step-3

$\exists C_r'$ such that

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

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

Direct L^r result: Step-3

$\exists C_r'$ such that

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

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

Direct L^r result: Step-3

$\exists C_r'$ such that

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

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

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

Direct L^r result: Step-3

$\exists C_r'$ such that

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

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

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

$$(*) \le C_r' m^{\frac{1}{\min(2,r)} = \max\{\frac{1}{2},\frac{1}{r}\}} \mathrm{vol}^{2/r}(\mathcal{K}).$$

Guarantee on derivatives with unbounded $supp(\Lambda)$

Assumptions:

1. $\mathbf{z} \mapsto \nabla_{\mathbf{z}}\left[\partial^{\mathbf{p},\mathbf{q}} k(\mathbf{z})\right]$: continuous; $\mathcal{K} \subset \mathbb{R}^d$: compact, $E_{\mathbf{p},\mathbf{q}} := \mathbb{E}_{\omega\sim\Lambda}|\boldsymbol{\omega}^{\mathbf{p}+\mathbf{q}}|\,\|\boldsymbol{\omega}\|_2 < \infty$.
2. $\exists L > 0, \sigma > 0$

$$\mathbb{E}_{\omega\sim\Lambda}|f(\mathbf{z};\boldsymbol{\omega})|^M \leq \frac{M!\,\sigma^2 L^{M-2}}{2} \quad (\forall M \geq 2, \forall \mathbf{z} \in \mathcal{K}_\Delta),$$

$$f(\mathbf{z};\boldsymbol{\omega}) = \partial^{\mathbf{p},\mathbf{q}} k(\mathbf{z}) - \boldsymbol{\omega}^{\mathbf{p}}(-\boldsymbol{\omega})^{\mathbf{q}} h_{|\mathbf{p}+\mathbf{q}|}(\boldsymbol{\omega}^T\mathbf{z}).$$

Guarantee on derivatives with unbounded $supp(\Lambda)$

Assumptions:

1. $\mathbf{z} \mapsto \nabla_{\mathbf{z}} \left[\partial^{\mathbf{p},\mathbf{q}} k(\mathbf{z})\right]$: continuous; $\mathcal{K} \subset \mathbb{R}^d$: compact, $E_{\mathbf{p},\mathbf{q}} := \mathbb{E}_{\boldsymbol{\omega}\sim\Lambda} |\boldsymbol{\omega}^{\mathbf{p}+\mathbf{q}}| \, \|\boldsymbol{\omega}\|_2 < \infty$.
2. $\exists L > 0, \sigma > 0$

$$\mathbb{E}_{\boldsymbol{\omega}\sim\Lambda} |f(\mathbf{z};\boldsymbol{\omega})|^M \leq \frac{M!\,\sigma^2 L^{M-2}}{2} \quad (\forall M \geq 2, \forall \mathbf{z} \in \mathcal{K}_\Delta),$$
$$f(\mathbf{z};\boldsymbol{\omega}) = \partial^{\mathbf{p},\mathbf{q}} k(\mathbf{z}) - \boldsymbol{\omega}^{\mathbf{p}}(-\boldsymbol{\omega})^{\mathbf{q}} h_{|\mathbf{p}+\mathbf{q}|}(\boldsymbol{\omega}^T \mathbf{z}).$$

Then with $F_d := d^{-\frac{d}{d+1}} + d^{\frac{1}{d+1}}$

$$\Lambda^m \left(\|\partial^{\mathbf{p},\mathbf{q}} k - s^{\mathbf{p},\mathbf{q}}\|_{L^\infty(\mathcal{K})} \geq \epsilon \right) \leq$$
$$\leq 2^{d-1} e^{-\frac{m\epsilon^2}{8\sigma^2\left(1+\frac{\epsilon L}{2\sigma^2}\right)}} + F_d 2^{\frac{4d-1}{d+1}} \left[\frac{|\mathcal{K}|(D_{\mathbf{p},\mathbf{q},\mathcal{K}} + E_{\mathbf{p},\mathbf{q}})}{\epsilon}\right]^{\frac{d}{d+1}} e^{-\frac{m\epsilon^2}{8(d+1)\sigma^2\left(1+\frac{\epsilon L}{2\sigma^2}\right)}},$$

where $D_{\mathbf{p},\mathbf{q},\mathcal{K}} := \sup_{\mathbf{z}\in conv(\mathcal{K}_\Delta)} \|\nabla_{\mathbf{z}} \left[\partial^{\mathbf{p},\mathbf{q}} k(\mathbf{z})\right]\|_2$.

Comments

- Proof idea: '[Rahimi and Recht, 2007]: Hoeffding (boundedness!) + Lipschitzness' → 'Bernstein + Lipschitzness'.

Comments

- Proof idea: '[Rahimi and Recht, 2007]: Hoeffding (boundedness!) + Lipschitzness' → 'Bernstein + Lipschitzness'.
- Example: Gaussian kernel.

Comments

- Proof idea: '[Rahimi and Recht, 2007]: Hoeffding (boundedness!) + Lipschitzness' → 'Bernstein + Lipschitzness'.
- Example: Gaussian kernel.
- It gives the (slightly worse)

$$\left\|\partial^{\mathbf{p},\mathbf{q}} k - s^{\mathbf{p},\mathbf{q}}\right\|_{L^\infty(\mathcal{K})} = O_{a.s.}\left(|\mathcal{K}|\sqrt{m^{-1}\log m}\right)$$

rate.

Summary

Finite sample

- $L^\infty(\mathcal{K})$ guarantees $\xrightarrow{\text{spec.}} |\mathcal{K}_m| = e^{o(m)}$ – optimal!
- $L^r(\mathcal{K})$ results ($\Leftarrow$ uniform, type of L^r).
- derivative approximation guarantees:
  - improved $|\mathcal{K}_m|$ growing – bounded spectral support.
  - handling unbounded spectral support.

Research directions

- Tighter derivative guarantees (unbounded empirical processes).
- Error propagation to prediction.
- LCA/Mercer, . . . extensions.

Thank you for the attention!

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.

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

Yurinsky, V. (1995).
Sums and Gaussian Vectors.
Springer.

Contents

- Borel-Cantelli lemma.
- McDiarmid inequality.
- Bernstein inequality.
- Support of a measure.
- $L^\infty(\mathcal{K})$ is *not* separable.

Borel-Cantelli lemma

- Assume: $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$.
- Then $\mathbb{P}(\infty\text{-ly many of them occur}) = 0$. 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 $\mathbb{P}\left(f(X_1, \ldots, X_m) < \mathbb{E}\left[f(X_1, \ldots, X_m)\right] + c\sqrt{\frac{\tau m}{2}}\right) \geq 1 - e^{-\tau}$.

Bernstein inequality [Yurinsky, 1995]

Let $(\xi_j)_{j=1}^m \overset{i.i.d.}{\sim} \mathbb{P}$, $\mathbb{E}_{\xi_j \sim \mathbb{P}}[\xi_j] = 0$, and assume that $\exists L > 0, S > 0$

$$\sum_{j=1}^m \mathbb{E}_{\xi_j \sim \mathbb{P}} \left[|\xi_j|^M \right] \leq \frac{M! S^2 L^{M-2}}{2} \quad (\forall M \geq 2).$$

Then for $\forall m \in \mathbb{N}^+$, $\forall \eta > 0$,

$$\mathbb{P}^m \left(\left| \sum_{j=1}^m \xi_j \right| \geq \eta S \right) \leq e^{-\frac{1}{2} \frac{\eta^2}{1 + \frac{\eta L}{S}}}.$$

Support of a measure

- Ingredients:
  - (X,τ): topological space with a countable basis.
  - $\mathcal{B} = \sigma(\tau)$: sigma-algebra generated by τ.
  - Λ: measure on $(X, \mathcal{B})$.

  Then

$$supp(\Lambda) = \overline{\cup\{A \in \tau : \Lambda(A) = 0\}},$$

  i.e., the complement of the union of all open Λ-null sets.
- Our choice: $X = \mathbb{R}^d$.

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

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