# Notes for Week 2/1/18 - 2/8/18

Alden Green

February 8, 2019

For fixed integers $n_1+n_2 = n$, let $\mathbf{X} = \{x_1, \ldots, x_{n_1}\} \subset \mathbb{R}^d$ and $\mathbf{Y} = \{y_1, \ldots, y_{n_2}\}$ be sampled i.i.d from distributions $\mathbb{P}$ and $\mathbb{Q}$ with density functions $p$ and $q$, respectively, both with support on $D \subset \mathbb{R}^d$. Our statistical problem is testing the null hypothesis $H_0 : \mathbb{P} = \mathbb{Q}$ vs. the alternative $H_1 : \mathbb{P} \neq \mathbb{Q}$, where our knowledge of $\mathbb{P}$ and $\mathbb{Q}$ come from the samples $\mathbf{X}$ and $\mathbf{Y}$.

Recall the *Laplacian smooth* and *total variation smooth* test statistics

$$T_1(\boldsymbol{\ell}; G_{n,r}) = \sup_{\theta : \|\mathbf{B}\theta\|_1 \leq C_{n,r}} \left|\boldsymbol{\ell}^T \theta\right|$$
$$T_2(\boldsymbol{\ell}; G_{n,r}) = \sup_{\theta : \|\mathbf{B}\theta\|_2 \leq C_{n,r}} \left|\boldsymbol{\ell}^T \theta\right|$$

where $\mathbf{B}$ is the incidence matrix of the $r$-neighborhood graph and $C_{n,r} = \frac{\sigma_k}{n^2 r_n^{d+2}}$.

Theorem 1 is the type of theorem we are looking for.

**Theorem 1.** *Under assumptions,*

$$\sqrt{n} T_2(\boldsymbol{\ell}; G_{n,r}) \rightsquigarrow ???$$

*under the null hypothesis* $H_0 : \mathbb{P} = \mathbb{Q}$.

The rest of this document details the strategy for proving this convergence.

## 1 Quantization

To ease proofs, we will assume $\mathcal{D} = [0,1]^d$.

Construct $G_{lat} = (V_{lat}, E_{lat})$ a lattice graph with equal side lengths in $[0,1]^d$, where

$$V_{lat} = P_{lat}(N) := \left\{ \left( \frac{i_1}{N} - \frac{1}{2N}, \cdots, \frac{i_d}{N} - \frac{1}{2N} \right) : i_1, \ldots, i_d \in \{1, \ldots, N\} \right\}$$
$$(z, z') \in E_{lat} \text{ if and only if } \|z - z'\| \leq \frac{1}{N}$$

where $z$ and $z' \in P_{lat}(N)$.

Denoting $I = P_{lat}$, we define

$$P_I(x) = \operatorname{argmin}\{\|x - z'\|_\infty, z' \in P_{lat}(N)\}$$

Then, let $C(z) = \{x \in [0,1]^d : z = P_I(x)\}$ be the collection of cells associated with the mesh $P_{lat}(N)$, noting that $\{C(z) : z \in P_{lat}(N)\}$ defines a partition over $[0,1]^d$.

For a given function $f$, the *quantization* $\bar{f}$ over $\{C(z) : z \in P_{lat}(N)\}$ is defined by

$$\overline{f}(x_0) = \frac{1}{\nu(C(z_0))} \int_{C(z_0)} f(x) d\nu(x)$$

where $x_0 \in C(z_0)$.

# 2 High-Level Proof Strategy for Theorem 1

For any function $f : \mathcal{D} \to \mathbb{R}$, let $\theta_f$ denote the evaluations of $f$ over $\mathbf{X}$, meaning

$$(\theta_f)_i = f(x_i)$$

(i) *Quantization:* Consider the function class

$$\mathcal{W}_n = \left\{ C_{n,r} \frac{\overline{f}}{\left\|\mathbf{B}\theta_{\bar{f}}\right\|_2} : f \in \mathcal{W}^{1,2}(\mathcal{D}, \rho^2) \right\}. \tag{1}$$

Under assumptions,

$$\sup_{\theta : \|B\theta\|_2 \le 1} \left|\boldsymbol{\ell}^T\theta\right| - \sup_{\widetilde{f} \in \mathcal{W}_n} \left|\mathbb{P}_n \widetilde{f} - \mathbb{Q}_n \widetilde{f}\right| = o_{\mathbb{P}}(n^{-1/2})$$

(ii) *Donsker-convergence of empirical process:* Write

$$\widetilde{f} = \frac{\overline{f}}{\left\|\mathbf{B}\theta_{\bar{f}}\right\|_2}.$$

Under assumptions,

$$\left\{ \mathbb{G}_{\mathbb{P}_n} \widetilde{f} : f \in \mathcal{W}^{1,2}(\mathcal{D}, \rho^2) \right\} \rightsquigarrow G_{\mathbb{P}}$$

and

$$\left\{ \mathbb{G}_{\mathbb{Q}_n} \widetilde{f} : f \in \mathcal{W}^{1,2}(\mathcal{D}, \rho^2) \right\} \rightsquigarrow G_{\mathbb{Q}},$$

where $G_{\mathbb{P}}$ is a tight Gaussian process with support on $\mathcal{W}^{1,2}(\mathcal{D}, \rho^2)$ and for measures $P_n$ and $P$, $\mathbb{G}_{P_n,P} = \sqrt{n}(P_n - P)$ and we suppress notational dependence on $P$ when obvious from context.

(iii) *Continuous mapping:* Under the null hypothesis $H_0 : \mathbb{P} = \mathbb{Q}$, for any $\widetilde{f} \in \mathcal{W}_n$,

$$\sqrt{n}(\mathbb{P}_n\widetilde{f} - \mathbb{Q}_n\widetilde{f}) = (\mathbb{G}_{\mathbb{P}_n} - \mathbb{G}_{\mathbb{Q}_n})\widetilde{f}$$

Thus, by i), the independence of $\mathbb{P}_n$ and $\mathbb{Q}_n$, and the continuous mapping theorem,

$$\left\{\sqrt{n}(\mathbb{P}_n\widetilde{f} - \mathbb{Q}_n\widetilde{f}), f \in \mathcal{W}^{1,2}(\mathcal{D}, \rho^2)\right\} \rightsquigarrow G_{\mathbb{P}} - G'_{\mathbb{P}}$$

where $G_{\mathbb{P}}$ and $G'_{\mathbb{P}}$ are i.i.d. Gaussian processes.

By the continuous mapping theorem again

$$\sup_{f \in \mathcal{W}^{1,2}(\mathcal{D},\rho^2)} \left|\sqrt{n}(\mathbb{P}_n\widetilde{f} - \mathbb{Q}_n\widetilde{f})\right| \rightsquigarrow \sup_{f \in \mathcal{W}^{1,2}(\mathcal{D},\rho^2)} |(G_{\mathbb{P}} - G'_{\mathbb{P}})f|$$

# 3 Proof Strategy for *Quantization*

For a given $\theta \in \mathbb{R}^n$, let the quantization of $\theta$ be given by

$$(\bar{\theta})_i = \frac{\sum_{x_j \in C(z_0)} \theta_j}{|C(z_0)|}$$

where $x_i \in C(z_0)$, and $|C(z_0)| = \sum_{j=1}^n \mathbf{1}(x_j \in C(z_0))$ is the number of data points in $C(z_0)$.

**Lemma 1.**

$$\sup_{\theta:\|B\theta\|_2 \le C_{n,r}} \left|\boldsymbol{\ell}^T\theta\right| \ge \sup_{\widetilde{f} \in \mathcal{W}_n} \left|\mathbb{P}_n\widetilde{f} - \mathbb{Q}_n\widetilde{f}\right|$$

*Proof.* For any $f : \mathcal{D} \to \mathbb{R}$, $\mathbb{P}_n f - \mathbb{Q}_n f = \boldsymbol{\ell}^T\theta_f$, and so

$$\sup_{\widetilde{f} \in \mathcal{W}_n} \left|\mathbb{P}_n\widetilde{f} - \mathbb{Q}_n\widetilde{f}\right| = \sup_{\theta_{\widetilde{f}}:\widetilde{f} \in \mathcal{W}_n} \left|\boldsymbol{\ell}^T\theta_{\widetilde{f}}\right|$$

Then, for $\widetilde{f} \in \mathcal{W}_n$

$$\left\|B\theta_{\widetilde{f}}\right\|_2 = \left\|B\theta_{\frac{C_{n,r}\bar{f}}{\left\|B\theta_{\bar{f}}\right\|_2}}\right\| = C_{n,r}.$$

and so $\left\{\theta_{\widetilde{f}} : \widetilde{f} \in \mathcal{W}_n\right\} \subset \{\theta : \|\mathbf{B}\theta\|_2 \le C_{n,r}\}$. □

**Lemma 2.** *For any $\theta \in \mathbb{R}^n$*

$$\boldsymbol{\ell}^T(\theta - \bar{\theta}) \le 2\|\mathbf{B}\theta\|_1$$

**Theorem 2.** *For any $\theta \in \mathbb{R}^n$ such that $\|\mathbf{B}\theta\|_2 \leq C_{n,r}$, there exists some $f \in \mathcal{W}^{1,2}(\mathcal{D}, \rho^2)$ such that*

$$\left|\mathbb{P}_n \widetilde{f} - \mathbb{Q}_n \widetilde{f}\right| \geq \left|\mathbb{P}_n \overline{\theta} - \mathbb{Q}_n \overline{\theta}\right| - o_{\mathbb{P}}(n^{-1/2})$$

*Proof.* For $\theta \in \mathbb{R}^n$, consider the quantized extension of $\theta$, $\overline{f}_\theta : \mathcal{D} \to \mathbb{R}$ defined by

$$\overline{f}_\theta(x) = (\overline{\theta})_i$$

for any $x_i$ in the same cell $C(z)$ as $x$. Clearly,

$$\mathbb{P}_n \overline{f}_\theta - \mathbb{Q}_n \overline{f}_\theta = \mathbb{P}_n \overline{\theta} - \mathbb{Q}_n \overline{\theta}$$

Now, we have to show that either $\overline{f}_\theta \in \mathcal{W}_n$, or that it can be well-approximated by some $\widetilde{f} \in \mathcal{W}_n$. $\square$

# 4 Proof strategy for *Donsker-convergence of empirical processes*:

Consider the function class $\mathcal{W}_n$ given by (1). We wish to show

(a) $\mathcal{W}_n$ is totally bounded.

(b) For every sequence $\delta_n \downarrow 0$,

$$\sup_{\|f-g\|_{1,2,\rho^2} \leq \delta_n} \mathbb{P}(\widetilde{f} - \widetilde{g})^2 \to 0$$

where $f$ and $g$ are arbitrary elements of $\mathcal{W}^{1,2}(\mathcal{D}, \rho^2)$.

(c) There exists a sequence of envelope functions $F_n$ on $\mathcal{W}_n$ satisfying the Lindeberg condition

$$\mathbb{P}F_n^2 = \mathcal{O}(1)$$
$$\mathbb{P}F_n^2 \mathbf{1} F_n > \epsilon\sqrt{n} \to 0 \qquad \text{(for any } \epsilon > 0\text{)}$$

(d) The bracketing integral

$$\int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{W}_n, L_2(\mathbb{P}))}$$

converges to 0 for any $\delta_n \downarrow 0$.

Then, the desired results holds by application of Theorem 3.

**Theorem 3** (Donsker theorem for changing function classes)**.** *Let $\mathcal{F}_n = \{f_{n,t} : t \in T\}$ be a class of measurable functions indexed by a totally bounded semimetric space $(T, \rho)$ satisfying*

$$\sup_{\rho(s,t)<\delta_n} P(f_{n,s} - f_{n,t})^2 \to 0, \quad \textit{every } \delta_n \downarrow 0$$

*and with envelope function $F_n$ satisfying the Lindeberg condition*

$$PF_n^2 = \mathcal{O}(1)$$
$$PF_n^2 \mathbf{1}\left\{F_n > \epsilon\sqrt{n}\right\} \to 0, \quad \textit{for every } \epsilon > 0.$$

*If $J_{[]}(\delta_n, \mathcal{F}_n, L_2(P)) \to 0$ for every $\delta_n \downarrow 0$, then the sequence $\{\mathbb{G}_n f_{n,t}, t \in T\}$ converges in distribution to a tight Gaussian process, provided the sequence of covariance functions*

$$Pf_{n,s}f_{n,t} - Pf_{n,s}Pf_{n,t}$$

*converges pointwise on $T \times T$.*