Recap: Variational Analysis

$\theta^* = \arg\min_\theta f(\theta)$ | θ^* : $f(\theta) = 0$

$\theta_n^* = \arg\min f_n(\theta)$ | θ_n^* : $f_n(\theta) = 0$

Q. If $f \approx f_n$ in some sense, how far θ and θ^* can be?
Can we quantify the proximity btw θ and θ^*?

Ex 1. LSE. $f(\theta) = \mathbb{E}(Y - X^T\theta)^2$ $\quad f_n(\theta) = \frac{1}{n}\sum_{i=1}^n (Y_i - X_i^T\theta)^2$

Ex 2. M estimation w/o constraints.

$$\theta^* = \arg\min_\theta \mathbb{E}\, \ell(X, \theta) \qquad \theta_n^* = \arg\min \frac{1}{n}\sum_{i=1}^n \ell(X_i, \theta)$$

(Hjort and Pollard, 2011) If $\ell(x, \cdot)$ is convex then
$|\theta^* - \theta_n^*| \le \varepsilon$ whenever ......

Ex 3. $\theta^* = \arg\min_\theta f(\theta)$ $\quad \theta_n^* = \arg\min_\theta f(\theta_0) + (\theta - \theta_0)^T \nabla f(\theta_0)$
$+ \frac{1}{2}(\theta - \theta_0)^T \nabla^2 f(\theta_0)(\theta - \theta_0)$

Kantorovich's theorem. Under some smoothness
$\|\theta^* - \theta_n^*\|_2 \lesssim \|\nabla_2 f(\theta_0)^{-1} \cdot \nabla f(\theta_0)\|_2^2$.

Def. Epi-convergence. For a seq. of $\overbrace{\text{fns}}^{\text{l.s.c.}}$ $\{f_n : \mathcal{X} \to \overline{\mathbb{R}}\}$ and a fn $f : \mathcal{X} \to \overline{\mathbb{R}}$
$f_n \xrightarrow{e} f$ if $\forall x \in \mathcal{X}$

① $\forall x_n \to x$ $\quad \liminf_n f_n(x_n) \ge f(x)$

② $\exists x_n \to x$ $\quad \limsup_n f_n(x_n) \le f(x)$

Rmk 1. $\theta^* = \underset{\theta \in C}{\operatorname{argmin}} f(\theta) = \operatorname{argmin} f(\theta) + \delta_C(\theta)$.

$$\delta_C(\theta) = \begin{cases} 0 & \theta \in C \\ \infty & \theta \notin C \end{cases}$$

θ^* activates C.

Rmk 2. We may consider $(\mathcal{X}, \|\cdot\|)$ or $(\mathbb{R}^d, \|\cdot\|)$

Thm 3.10 of Attouch and Wets (1983)

Let $x_n \in \operatorname{argmin} f_n(x)$, $x_n \longrightarrow x$, and $f_n \xrightarrow{e} f$ then

$x \in \operatorname{argmin} f(x)$.

pf. $\forall y \in \mathcal{X}$ $\exists y_n \to y$ s.t $\limsup_n f_n(y_n) \le f(y)$.

$f(x) \le \liminf_n f_n(x_n) \le \liminf_n f_n(y_n) \le \limsup_n f_n(y_n) \le f(y)$

Epi-convergence $\Rightarrow$ minimizers converge.

1. Stronger result is possible.
2. whether $\exists\, d$ on lsc-fn(X) s.t $d(f_n, f) \to 0 \Rightarrow f_n \xrightarrow{e} f$

whether $\exists\, F$ s.t $d(x_n, x) \le F(d(f_n, f))$.

Painleve - Kuratowski convergence (PK-convergence)

Def. Painleve-Kuratowski (PK) convergence of sets $\{A_n \in \mathcal{X}\}$.

$$\limsup_n A_n = \{x \in \mathcal{X} : \liminf_n d(x, A_n) = 0\}$$
$$= \{x \in \mathcal{X} : \exists \{x_{n_k} \in A_{n_k}\} \text{ s.t. } x_{n_k} \to x\}$$

$$\liminf_n A_n = \{x \in \mathcal{X} : \limsup_n d(x, A_n) = 0\}$$
$$= \{x \in \mathcal{X} : \exists \{x_n \in A_n\} \text{ s.t. } x_n \to x\}$$

$\text{PK-}\lim A_n := \limsup_n A_n \overset{①}{=} \liminf_n A_n$ if ① hold

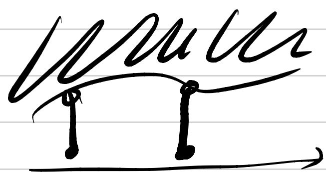
① For a seq of closed constraints $\{C_n\}$

$$\delta_{C_n} \xrightarrow{epi} \delta_C \iff \text{PK-}\lim C_n = C.$$

②

Def. Epi-graph.. For a lsc ftn f define

$$\text{epi-}f = \{(x, \alpha) \in \mathcal{X} \times \mathbb{R} : f(x) \leq \alpha\}$$



$f_n \xrightarrow{e} f$ iff

$$\text{PK-}\lim \text{epi-}f_n = \text{epi-}f.$$

Useful in quantifying a distance btw f_n & f

$$\mathcal{d}(f_n, f) = \text{Haus}(\text{epi-}f_n, \text{epi-}f)$$

Thm 3.8 of Rockafeller and Wets (2018)

$\forall \varepsilon_n \downarrow 0$ $\limsup_n \varepsilon_n\text{-}\arg\min f_n \subseteq \arg\min f$.

$\varepsilon_n\text{-}\arg\min f_n = \{\theta : f_n(\theta) \leq \inf f_n(\theta) + \varepsilon_n\}$.

Application 1. Unconstrained M estimation

$$\theta^* = \arg\min M(\theta) = \arg\min \mathbb{E}\, \ell(X, \theta)$$

$$\hat{\theta}_n \in \arg\min \hat{M}_n(\theta) = \arg\min \tfrac{1}{n} \textstyle\sum_{i=1}^n \ell(X_i, \theta)$$

If $\hat{M}_n(\theta) \xrightarrow{e} M(\theta)$ a.s., then $\forall \varepsilon_n \downarrow 0$. a.s.

$$\limsup_n \underbrace{\{\theta : \hat{M}_n(\theta) \leq \hat{M}_n(\hat{\theta}_n) + \varepsilon_n\}}_{\hat{CI}_n} \subseteq \{\theta_0\}.$$

① $\varepsilon_n = 0$ → a.s. convergence of M-estimator.

② If $\hat{CI}_n$ is an (asymptotically) valid confidence set for θ_0,

$\text{Diam}(\hat{CI}_n) \to 0$. (Takatsu and Kuchibhotla 2025+)

set convergence in stat

Section 7 Royset (2020), Rockafellar and Royset (2024), Aswani (2019)

Application 2. Constrained M-estimation.

$$\hat{\theta}_n = \operatorname*{argmin}_{\theta \in \Theta_n} \hat{M}_n(\theta) = \operatorname{argmin}_\theta \hat{M}_n(\theta) + \delta_{\Theta_n}(\theta)$$

$$r_n(\hat{\theta}_n - \theta_0) = \operatorname*{argmin}_u s_n \{ \hat{M}_n(\theta_0 + \tfrac{u}{r_n}) - \hat{M}_n(\theta_0) \}$$

$$+ \delta_{\Theta_n}(\theta_0 + \tfrac{u}{r_n}) \quad \delta_{r_n(\Theta_n - \theta_0)}(u)$$

$$= \operatorname{argmin} \; s_n \{ (\hat{M}_n - M)(\theta_0 + \tfrac{u}{r_n}) - (\hat{M}_n - M)(\theta_0) \} \xrightarrow{e} S(u)$$

Pollard 1984 Brownian Motion.

$$+ \; s_n \{ M(\theta_0 + \tfrac{u}{r_n}) - M(\theta_0) \} \xrightarrow{e} D(u)$$

$$+ \; \delta_{r_n(\Theta_n - \theta_0)}(u) \xrightarrow{e} \delta_C(u)$$

Then, $r_n(\hat{\theta}_n - \theta_0) \xrightarrow{d} \operatorname{argmin} S(u) + D(u) + \delta_C(u)$

① $f_n \xrightarrow{e} f$ and $g_n \xrightarrow{e} g \Rightarrow f_n + g_n \xrightarrow{e} f + g$ (*)

(*) is not true in general.

Prop 3.4 of Royset (2020)

If either of the following holds, then (*) holds.

① $-f_n \xrightarrow{e} -f$ or $-g_n \xrightarrow{e} -g$.

② $f_n \to f$ and $g_n \to g$ pointwise.

① $\iff \lim_n f_n(x_n) = f(x)$. $\forall x_n \to x$.

② For a stochastic process. Skorokhod. representation

a.s. convergence $\iff$ convergence in dist.

a.s. epi-convergence $\iff$ epi-convergence in dist.

(A.s. representations ~~tof~~ stochastic Process.

Thm 2.3 of Kim and Pollard, Thm 1.10.4 of van der vaart and Wellner.

For a stochastic process $\{f_n\}$ defined on $\mathbb{R}^d$

① $(f_n(t_1), \cdots f_n(t_m))^T \xrightarrow{d} g(t_1 \cdots t_m)$

② $\{f_n\}$ is uniformly tight.

Then $\exists$ prob space on where $\exists \tilde{f}_n, \tilde{f}$ s.t.

$f_n \overset{d}{=} \tilde{f}_n$. $(\tilde{f}(t_1), \cdots \tilde{f}(t_m)) \overset{d}{=} g(t_1 \cdots t_m)$ and

$\sup_{u \in K} |\tilde{f}_n(u) - \tilde{f}(u)| \to 0$ a.s.

Thm 1 of Knight (1999)

If $Z_n \xrightarrow{e-d} Z$, $\varepsilon_n \xrightarrow{P} 0$, $U_n \in \varepsilon_n$- argmin Z_n. and $U_n = O_p(1)$

then $U_n \xrightarrow{d} U := \operatorname{argmin} Z$.

Tangent cones. One can verify they are closed cones

Contingent cone. $T_C(u) = \limsup_{\tau \downarrow 0} \frac{C-u}{\tau} = \{ v : \liminf_{\tau \downarrow 0} \frac{d(u+\tau v, C)}{\tau} = 0 \}$

Adjacent cone $\tilde{T}_C(u) = \{ v : \limsup_{\tau \downarrow 0} \frac{d(u+\tau v, C)}{\tau} = 0 \}$

Clarke cone $T_C^{\dagger}(u) = \{ v : \limsup_{\substack{\tau \downarrow 0 \\ u' \to u}} \frac{d(u'+\tau v, C)}{\tau} = 0 \}$
convex

$T_C^{\dagger}(u) \subseteq \tilde{T}_C(u) \subseteq T_C(u)$ $u \in C^{\circ}$ then they are $\mathbb{R}^d$.

If C is closed and convex, then, they are all same
(Clarke, 1983)

C is Chernoff regular at u if $\tilde{T}_C(u) = T_C(u)$
(Chernoff 1954, Geyer 1994) is used for asymptotic for global minimizer (Geyer 1994)

C is Clarke regular at u if $T_C^{\dagger}(u) = T_C(u)$

(Geyer 1994) claimed if Clarke regular at true parameter local minimizer. ; they are asym. equivalent
one can identify asym. dist. of

(Shapiro, 2000) This is not true, stronger assump. for

Application. Constraint mean-estimation.

Ex. $\hat{\theta}_n = \operatorname{argmin} \frac{1}{n} \sum \|X_i - \theta\|^2 = \operatorname*{argmin}_{\theta \in \Theta} \| \bar{X}_n - \theta \|^2$

$= \mathrm{Proj}_\Theta (\bar{X}_n)$

$$\| \mathrm{Proj}_\Theta(\bar{X}_n) - \underbrace{\mathrm{Proj}_\Theta(\mu)}^{\theta_0} \| \le \| \bar{X}_n - \mu \| \xrightarrow{a.s.} 0$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \operatorname*{argmin}_u s_n \| \bar{X}_n - \theta_0 - \frac{u}{\sqrt{n}} \|^2$$

$$= '' - \frac{s_n}{\sqrt{n}} 2(\bar{X}_n - \theta_0)^T u + \frac{s_n}{n} \|u\|^2 + \delta_{\sqrt{n}(\Theta - \theta_0)}(u)$$

$$\xrightarrow{\operatorname{argmin}} -2Z^T u + \|u\|^2 = Z$$

clarke regular at $x \in [0,1]$

$\Theta = [0,1] \to T_C(x) = \mathbb{R} \quad \forall x \in (0,1)$

$T_C(0) = [0, \infty)$ $\qquad Z$

$T_C(1) = (-\infty, 0] \to \max\{Z, 0\}$

$\min\{Z, 0\}$