学習理論特論 (#02) 補足資料

包 含

2026 年 10 月 20 日

今回の講義中では，積分と inf の入れ替えを伴う操作が発生する．時間の都合上，講義中ではその詳細をカバーすることができないため，ここで補足を行う．

■設定． 損失関数 $\phi\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ は Borel 可測とする．ここでは凸性どころか連続性も要請しない．従って，ロジスティック損失やヒンジ損失のような凸損失だけでなく，01 損失のような非連続な損失関数も範疇に入っている．この損失関数に対して，期待リスク $R_\phi(f) \coloneqq \mathbb{E}_{(X,Y)}[\phi(Yf(X))]$ を可測関数である分類器 $f : \mathcal{X} \to \mathbb{R}$ について考える．変分解析を行うため，期待リスクの infimum である Bayes リスク $R^*_\phi \coloneqq \inf_{f\in\mathcal{M}(\mathcal{X},\mathbb{R})} R_\phi(f)$，および各点で定義される量

- 条件付きリスク $C_{\phi,\eta}(\alpha) \coloneqq \eta\phi(\alpha) + (1-\eta)\phi(-\alpha)$
- 条件付き Bayes リスク $C^*_{\phi,\eta} \coloneqq \inf_{\alpha\in\mathbb{R}} C_{\phi,\eta}(\alpha)$

を考える．ただし $\mathcal{M}(\mathcal{X},\mathbb{R})$ は $\mathbb{R}^{\mathcal{X}}$ のうち可測な関数をすべて集めた集合である．

■主張． 以下，η として $\eta(X) = \Pr(Y = 1|X)$ (a.s.) を満たす可測関数 $\eta : \mathcal{X} \to [0,1]$ をとり，$\eta(X)$ をしばしば省略して η と書く．すると期待値と inf の入れ替えが成立する：

$$R^*_\phi = \inf_{f\in\mathcal{M}(\mathcal{X},\mathbb{R})} \mathop{\mathbb{E}}_{(X,Y)}[\phi(Yf(X))] = \mathop{\mathbb{E}}_{X}\left[\inf_{\alpha\in\mathbb{R}} C_{\phi,\eta}(\alpha)\right] = \mathop{\mathbb{E}}_{X}[C^*_{\phi,\eta}].$$

■証明の大筋． 入れ替えの等号のうち，$R^*_\phi \geq \mathbb{E}[C^*_{\phi,\eta}]$ の向きは自明に従う．従って，我々が示すべき方向は $R^*_\phi \leq \mathbb{E}[C^*_{\phi,\eta}]$ である．このために，各点で $C_{\phi,\eta}(f_\varepsilon(x)) < C^*_{\phi,\eta} + \varepsilon$ が成り立つ f_ε，すなわち pointwise minimizer をまず構成する．これが構成できれば，条件付きリスクの両辺 $\mathbb{E}_X$ で「期待値を取れば」$R_\phi(f_\varepsilon) < R^*_\phi + \varepsilon$ であり，$\varepsilon \to 0$ の極限を取ることで所望の方向が示せる．問題は「期待値を取れば」の部分であり，pointwise minimizer が可測であるかどうかが自明でない．証明の本質は f_ε **がきちんと可測になっている（すなわち可測選択になっている）**事実を示すことである．これには実は空間の可分性が効いてくる．

■証明． 証明の前半では，条件付きリスクの各点最小化は（一般には非可算 $\alpha \in \mathbb{R}$ 上で定義されているが）可算空間上で行ってもよいことを確認する．まず $S \coloneqq \{(\phi(\alpha), \phi(-\alpha)) \mid \alpha \in \mathbb{R}\} (\subseteq \mathbb{R}^2)$ を取る．これは可分空間であるから，S 中で稠密な部分集合 $D \coloneqq \{(\phi(\alpha_k), \phi(-\alpha_k))\}_{k\in\mathbb{N}}$ を取ることができる．$(\alpha_k)_{k\in\mathbb{N}}$ は可算列である．いま，任意の $\eta \in [0,1]$ に対して以下の評価が成立：

$$|C_{\phi,\eta}(\alpha) - C_{\phi,\eta}(\alpha')| \leq \max\{|\phi(\alpha) - \phi(\alpha')|, |\phi(-\alpha) - \phi(-\alpha')|\} \quad \forall \alpha, \alpha' \in \mathbb{R}. \tag{1}$$

これを用いると，任意の $\delta > 0$ について以下が従う：

$$\exists \alpha \in \mathbb{R}, \quad C_{\phi,\eta}(\alpha_k) \leq C_{\phi,\eta}(\alpha) + \delta < C^*_{\phi,\eta} + 2\delta. \tag{2}$$

2 つ目の不等号は $C^*_{\phi,\eta}$ の inf の定義から従う．一方，1 つ目の不等号は D の稠密性から従う．つまり，点 $(\phi(\alpha), \phi(-\alpha)) \in S$ にいくらでも近い点 $(\phi(\alpha_k), \phi(-\alpha_k)) \in D$ を取ることができ，ある誤差量 δ を取れば必ずその誤差量を守る k が存在する．なお，本来は S や D は $\mathbb{R}^2$ の部分集合なのでユークリッド距離を用いて点間距離を測るのが自然だが，$\mathbb{R}^2$ 内では任意のノルムが同値になるため，D は max ノルムに関しても S 内で稠密である．そして式 (1) より $|C(\alpha) - C(\alpha')|$ は max ノルムの距離で上から抑えられている．式 (2) より，$\delta \to 0$ の極限を取れば

$$C^*_{\phi,\eta} = \inf_{k \in \mathbb{N}} C_{\phi,\eta}(\alpha_k) \tag{3}$$

である（反対側の不等号は容易）．すなわち，条件付きリスクの最小化は非可算濃度の空間 $\mathbb{R}$ 内で行うかわりに，可算濃度の空間内で行っても同じであることが確認できた．

ここからの後半では，実際に pointwise minimizer が可測関数であることを見ていく．仮定より $x \mapsto C_{\phi,\eta(x)}(\alpha_k)$ は可測（各 k を止めたときに）なので，同様に $x \mapsto C^*_{\phi,\eta(x)}$ も可測である．これより，$\varepsilon > 0$ を固定したとき

$$k(x) \coloneqq \min \left\{ k \in \mathbb{N} \;\middle|\; C_{\phi,\eta(x)}(\alpha_k) < C^*_{\phi,\eta(x)} + \varepsilon \right\}$$

と定め，同時に $f_\varepsilon(x) \coloneqq \alpha_{k(x)}$ と定める．$k(x) < \infty$ は ϕ の非負性から直ちにわかる．また，$\{x \in \mathcal{X} \mid k(x) = k\}$ は有限個の可測集合の共通部分で書けるので，f_ε はその値域が可算集合に含まれるような関数になっている．ここから f_ε の可測性も従う．なぜなら，任意の可測集合 B に対して $f_\varepsilon^{-1}(B) = \bigcup_{k:\alpha_k \in B} \{x \mid k(x) = k\}$ と書くことができ，右辺は可算和であることから f_ε の可測性が確認できる．従って，

$$(R^*_\phi \le)\; R_\phi(f_\varepsilon) = \mathop{\mathbb{E}}_X \left[C_{\phi,\eta(X)}(f_\varepsilon(X)) \right] < \mathop{\mathbb{E}}_X \left[C^*_{\phi,\eta(X)} \right] + \varepsilon = R^*_\phi + \varepsilon$$

の方向の不等式が成り立つ．$\varepsilon \to 0$ ととれば $R^*_\phi = \mathbb{E}[C^*_{\phi,\eta}]$，すなわち期待値と infimum の可換性が示された． □

■補足． このタイプの期待値と infimum の入れ替えは変分解析でしばしば現れる．より一般には Rockafellar-Wets (1998) “Variational Analysis” (Springer) の 14 章によくまとまっている．特に，定理 14.60 は被積分関数が「normal integrand」と呼ばれるクラスのものであれば，期待値と infimum が可換であることを主張している．ここでは凸損失は normal integrand だが，01 損失が normal integrand ではないため，より初等的な測度論に従って示した．