Appendix: *Classification Calibration Dimension for General Multiclass Losses*

Calculation of Trigger Probability Sets for Figure 2

(a) 0-1 loss $\ell^{\text{0-1}}$ ($n = 3$).

$$\boldsymbol{\ell}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} ; \ \boldsymbol{\ell}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} ; \ \boldsymbol{\ell}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} .$$

$$\begin{aligned}
\mathcal{Q}_1^{\text{0-1}} &= \{\mathbf{p} \in \Delta_3 : \mathbf{p}^\top \boldsymbol{\ell}_1 \le \mathbf{p}^\top \boldsymbol{\ell}_2, \ \mathbf{p}^\top \boldsymbol{\ell}_1 \le \mathbf{p}^\top \boldsymbol{\ell}_3\} \\
&= \{\mathbf{p} \in \Delta_3 : p_2 + p_3 \le p_1 + p_3, \ p_2 + p_3 \le p_1 + p_2\} \\
&= \{\mathbf{p} \in \Delta_3 : p_2 \le p_1, \ p_3 \le p_1\} \\
&= \{\mathbf{p} \in \Delta_3 : p_1 \ge \max(p_2, p_3)\}
\end{aligned}$$

By symmetry,

$$\begin{aligned}
\mathcal{Q}_2^{\text{0-1}} &= \{\mathbf{p} \in \Delta_3 : p_2 \ge \max(p_1, p_3)\} \\
\mathcal{Q}_3^{\text{0-1}} &= \{\mathbf{p} \in \Delta_3 : p_3 \ge \max(p_1, p_2)\}
\end{aligned}$$

(b) Ordinal regression loss ℓ^{ord} ($n = 3$).

$$\boldsymbol{\ell}_1 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} ; \ \boldsymbol{\ell}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} ; \ \boldsymbol{\ell}_3 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} .$$

$$\begin{aligned}
\mathcal{Q}_1^{\text{ord}} &= \{\mathbf{p} \in \Delta_3 : \mathbf{p}^\top \boldsymbol{\ell}_1 \le \mathbf{p}^\top \boldsymbol{\ell}_2, \ \mathbf{p}^\top \boldsymbol{\ell}_1 \le \mathbf{p}^\top \boldsymbol{\ell}_3\} \\
&= \{\mathbf{p} \in \Delta_3 : p_2 + 2p_3 \le p_1 + p_3, \ p_2 + 2p_3 \le 2p_1 + p_2\} \\
&= \{\mathbf{p} \in \Delta_3 : p_2 + p_3 \le p_1, \ p_3 \le p_1\} \\
&= \{\mathbf{p} \in \Delta_3 : 1 - p_1 \le p_1\} \\
&= \{\mathbf{p} \in \Delta_3 : p_1 \ge \tfrac{1}{2}\}
\end{aligned}$$

By symmetry,

$$\mathcal{Q}_3^{\text{ord}} = \{\mathbf{p} \in \Delta_3 : p_3 \ge \tfrac{1}{2}\}$$

Finally,

$$\begin{aligned}
\mathcal{Q}_2^{\text{ord}} &= \{\mathbf{p} \in \Delta_3 : \mathbf{p}^\top \boldsymbol{\ell}_2 \le \mathbf{p}^\top \boldsymbol{\ell}_1, \ \mathbf{p}^\top \boldsymbol{\ell}_2 \le \mathbf{p}^\top \boldsymbol{\ell}_3\} \\
&= \{\mathbf{p} \in \Delta_3 : p_1 + p_3 \le p_2 + 2p_3, \ p_1 + p_3 \le 2p_1 + p_2\} \\
&= \{\mathbf{p} \in \Delta_3 : p_1 \le p_2 + p_3, \ p_3 \le p_1 + p_2\} \\
&= \{\mathbf{p} \in \Delta_3 : p_1 \le 1 - p_1, \ p_3 \le 1 - p_3\} \\
&= \{\mathbf{p} \in \Delta_3 : p_1 \le \tfrac{1}{2}, \ p_3 \le \tfrac{1}{2}\}
\end{aligned}$$

(c) 'Abstain' loss $\ell^{(?)}$ ($n = 3$).

$$\boldsymbol{\ell}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} ; \ \boldsymbol{\ell}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} ; \ \boldsymbol{\ell}_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} ; \ \boldsymbol{\ell}_4 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} .$$

$$\begin{aligned}
\mathcal{Q}_1^{(?)} &= \{\mathbf{p} \in \Delta_3 : \mathbf{p}^\top \boldsymbol{\ell}_1 \le \mathbf{p}^\top \boldsymbol{\ell}_2, \ \mathbf{p}^\top \boldsymbol{\ell}_1 \le \mathbf{p}^\top \boldsymbol{\ell}_3, \ \mathbf{p}^\top \boldsymbol{\ell}_1 \le \mathbf{p}^\top \boldsymbol{\ell}_4\} \\
&= \{\mathbf{p} \in \Delta_3 : p_2 + p_3 \le p_1 + p_3, \ p_2 + p_3 \le p_1 + p_2, \ p_2 + p_3 \le \tfrac{1}{2}(p_1 + p_2 + p_3)\} \\
&= \{\mathbf{p} \in \Delta_3 : p_2 \le p_1, \ p_3 \le p_1, \ p_2 + p_3 \le \tfrac{1}{2}\} \\
&= \{\mathbf{p} \in \Delta_3 : p_1 \ge \tfrac{1}{2}\}
\end{aligned}$$

By symmetry,

$$\begin{aligned}\mathcal{Q}_2^{(?)} &= \{\mathbf{p} \in \Delta_3 : p_2 \geq \tfrac{1}{2}\} \\ \mathcal{Q}_3^{(?)} &= \{\mathbf{p} \in \Delta_3 : p_3 \geq \tfrac{1}{2}\}\end{aligned}$$

Finally,

$$\begin{aligned}\mathcal{Q}_4^{(?)} &= \{\mathbf{p} \in \Delta_3 : \mathbf{p}^\top \boldsymbol{\ell}_4 \leq \mathbf{p}^\top \boldsymbol{\ell}_1,\ \mathbf{p}^\top \boldsymbol{\ell}_4 \leq \mathbf{p}^\top \boldsymbol{\ell}_2,\ \mathbf{p}^\top \boldsymbol{\ell}_4 \leq \mathbf{p}^\top \boldsymbol{\ell}_2\} \\ &= \{\mathbf{p} \in \Delta_3 : \tfrac{1}{2}(p_1 + p_2 + p_3) \leq \min(p_2 + p_3, p_1 + p_3, p_1 + p_2)\} \\ &= \{\mathbf{p} \in \Delta_3 : \tfrac{1}{2} \leq 1 - \max(p_1, p_2, p_3)\} \\ &= \{\mathbf{p} \in \Delta_3 : \max(p_1, p_2, p_3) \leq \tfrac{1}{2}\}\end{aligned}$$

Proof of Theorem 6

Proof. Since ψ is classification calibrated w.r.t. ℓ over Δ_n, by Lemma 2, $\exists \text{pred}' : S_\psi \to [k]$ such that

$$\forall \mathbf{p} \in \Delta_n : \quad \inf_{\mathbf{z}' \in \mathcal{S}_\psi : \text{pred}'(\mathbf{z}') \notin \text{argmin}_t \mathbf{p}^\top \boldsymbol{\ell}_t} \mathbf{p}^\top \mathbf{z}' \ > \ \inf_{\mathbf{z}' \in \mathcal{S}_\psi} \mathbf{p}^\top \mathbf{z}' \,. \tag{9}$$

Now suppose there is some $\mathbf{z} \in \mathcal{S}_\psi$ such that $\mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z})$ is not contained in $\mathcal{Q}_t^\ell$ for any $t \in [k]$. Then $\forall t \in [k]$, $\exists \mathbf{q} \in \mathcal{N}_{S_\psi}(\mathbf{z})$ such that $\mathbf{q} \notin \mathcal{Q}_t^\ell$, i.e. such that $t \notin \text{argmin}_{t'} \mathbf{q}^\top \boldsymbol{\ell}_{t'}$. In particular, for $t = \text{pred}'(\mathbf{z})$, $\exists \mathbf{q} \in \mathcal{N}_{S_\psi}(\mathbf{z})$ such that $\text{pred}'(\mathbf{z}) \notin \text{argmin}_{t'} \mathbf{q}^\top \boldsymbol{\ell}_{t'}$.

Since $\mathbf{q} \in \mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z})$, we have

$$\mathbf{q}^\top \mathbf{z} \ = \ \inf_{\mathbf{z}' \in \mathcal{S}_\psi} \mathbf{q}^\top \mathbf{z}' \,. \tag{10}$$

Moreover, since $\text{pred}'(\mathbf{z}) \notin \text{argmin}_{t'} \mathbf{q}^\top \boldsymbol{\ell}_{t'}$, we have

$$\inf_{\mathbf{z}' \in \mathcal{S}_\psi : \text{pred}'(\mathbf{z}') \notin \text{argmin}_{t'} \mathbf{q}^\top \boldsymbol{\ell}_{t'}} \mathbf{q}^\top \mathbf{z}' \ \leq \ \mathbf{q}^\top \mathbf{z} \ = \ \inf_{\mathbf{z}' \in \mathcal{S}_\psi} \mathbf{q}^\top \mathbf{z}' \,. \tag{11}$$

This contradicts Eq. (9). Thus it must be the case that $\forall \mathbf{z} \in \mathcal{S}_\psi$, $\exists t \in [k]$ with $\mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z}) \subseteq \mathcal{Q}_t^\ell$. □

Proof of Theorem 7

The proof uses the following technical lemma:

Lemma 15. *Let $\psi : [n] \times \widehat{\mathcal{T}} \to \mathbb{R}_+$. Suppose there exist $r \in \mathbb{N}$ and $\mathbf{z}_1, \ldots, \mathbf{z}_r \in \mathcal{R}_\psi$ such that $\bigcup_{j=1}^r \mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z}_j) = \Delta_n$. Then any element $\mathbf{z} \in \mathcal{S}_\psi$ can be written as $\mathbf{z} = \mathbf{z}' + \mathbf{z}''$ for some $\mathbf{z}' \in \text{conv}(\{\mathbf{z}_1, \ldots, \mathbf{z}_r\})$ and $\mathbf{z}'' \in \mathbb{R}_+^n$.*

Proof. Let $\mathcal{S}' = \{\mathbf{z}' + \mathbf{z}'' : \mathbf{z}' \in \text{conv}(\{\mathbf{z}_1, \ldots, \mathbf{z}_r\}), \mathbf{z}'' \in \mathbb{R}_+^n\}$, and suppose there exists a point $\mathbf{z} \in \mathcal{S}_\psi$ which cannot be decomposed as claimed, i.e. such that $\mathbf{z} \notin \mathcal{S}'$. Then by the Hahn-Banach theorem (e.g. see [19], corollary 3.10), there exists a hyperplane that strictly separates $\mathbf{z}$ from $\mathcal{S}'$, i.e. $\exists \mathbf{w} \in \mathbb{R}^n$ such that $\mathbf{w}^\top \mathbf{z} < \mathbf{w}^\top \mathbf{a}\ \forall \mathbf{a} \in \mathcal{S}'$. It is easy to see that $\mathbf{w} \in \mathbb{R}_+^n$ (since a negative component in $\mathbf{w}$ would allow us to choose an element $\mathbf{a}$ from $\mathcal{S}'$ with arbitrarily small $\mathbf{w}^\top \mathbf{a}$).

Now consider the vector $\mathbf{q} = \mathbf{w} / \sum_{i=1}^n w_i \in \Delta_n$. Since $\bigcup_{j=1}^r \mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z}_j) = \Delta_n$, $\exists j \in [r]$ such that $\mathbf{q} \in \mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z}_j)$. By definition of positive normals, this gives $\mathbf{q}^\top \mathbf{z}_j \leq \mathbf{q}^\top \mathbf{z}$, and therefore $\mathbf{w}^\top \mathbf{z}_j \leq \mathbf{w}^\top \mathbf{z}$. But this contradicts our construction of $\mathbf{w}$ (since $\mathbf{z}_j \in \mathcal{S}'$). Thus it must be the case that every $\mathbf{z} \in \mathcal{S}_\psi$ is also an element of $\mathcal{S}'$. □

Proof. (Proof of Theorem 7)

We will show classification calibration of ψ w.r.t. ℓ (over Δ_n) via Lemma 2. For each $j \in [r]$, let

$$T_j = \left\{ t \in [k] : \mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z}_j) \subseteq \mathcal{Q}_t^\ell \right\};$$

by assumption, $T_j \neq \emptyset \; \forall j \in [r]$. By Lemma 15, for every $\mathbf{z} \in \mathcal{S}_\psi$, $\exists \boldsymbol{\alpha} \in \Delta_r, \mathbf{u} \in \mathbb{R}^n_+$ such that $\mathbf{z} = \sum_{j=1}^r \alpha_j \mathbf{z}_j + \mathbf{u}$. For each $\mathbf{z} \in \mathcal{S}_\psi$, arbitrarily fix a unique $\boldsymbol{\alpha}^\mathbf{z} \in \Delta_r$ and $\mathbf{u}^\mathbf{z} \in \mathbb{R}^n_+$ satisfying the above, i.e. such that

$$\mathbf{z} = \sum_{j=1}^r \alpha_j^\mathbf{z} \mathbf{z}_j + \mathbf{u}^\mathbf{z} .$$

Now define $\text{pred}' : \mathcal{S}_\psi \to [k]$ as

$$\text{pred}'(\mathbf{z}) = \min \left\{ t \in [k] : \exists j \in [r] \text{ such that } \alpha_j^\mathbf{z} \geq \tfrac{1}{r} \text{ and } t \in T_j \right\} .$$

We will show pred' satisfies the condition for classification calibration.

Fix any $\mathbf{p} \in \Delta_n$. Let

$$J_\mathbf{p} = \Big\{ j \in [r] : \mathbf{p} \in \mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z}_j) \Big\} ;$$

since $\Delta_n = \bigcup_{j=1}^r \mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z}_j)$, we have $J_\mathbf{p} \neq \emptyset$. Clearly,

$$\forall j \in J_\mathbf{p} : \mathbf{p}^\top \mathbf{z}_j = \inf_{\mathbf{z} \in \mathcal{S}_\psi} \mathbf{p}^\top \mathbf{z} \tag{12}$$

$$\forall j \notin J_\mathbf{p} : \mathbf{p}^\top \mathbf{z}_j > \inf_{\mathbf{z} \in \mathcal{S}_\psi} \mathbf{p}^\top \mathbf{z} \tag{13}$$

Moreover, from definition of T_j, we have

$$\forall j \in J_\mathbf{p} : \quad t \in T_j \implies \mathbf{p} \in \mathcal{Q}_t^\ell \implies t \in \text{argmin}_{t'} \mathbf{p}^\top \boldsymbol{\ell}_{t'} .$$

Thus we get

$$\forall j \in J_\mathbf{p} : \quad T_j \subseteq \text{argmin}_{t'} \mathbf{p}^\top \boldsymbol{\ell}_{t'} . \tag{14}$$

Now, for any $\mathbf{z} \in \mathcal{S}_\psi$ for which $\text{pred}'(\mathbf{z}) \notin \arg\min_{t'} \mathbf{p}^\top \boldsymbol{\ell}_{t'}$, we must have $\alpha_j^\mathbf{z} \geq \frac{1}{r}$ for at least one $j \notin J_\mathbf{p}$ (otherwise, we would have $\text{pred}'(\mathbf{z}) \in T_j$ for some $j \in J_\mathbf{p}$, giving $\text{pred}'(\mathbf{z}) \in \arg\min_{t'} \mathbf{p}^\top \boldsymbol{\ell}_{t'}$, a contradiction). Thus we have

$$\begin{aligned}
\inf_{\mathbf{z} \in \mathcal{S}_\psi : \text{pred}'(\mathbf{z}) \notin \text{argmin}_{t'} \mathbf{p}^\top \boldsymbol{\ell}_{t'}} \mathbf{p}^\top \mathbf{z} &= \inf_{\mathbf{z} \in \mathcal{S}_\psi : \text{pred}'(\mathbf{z}) \notin \text{argmin}_{t'} \mathbf{p}^\top \boldsymbol{\ell}_{t'}} \sum_{j=1}^r \alpha_j^\mathbf{z} \mathbf{p}^\top \mathbf{z}_j + \mathbf{p}^\top \mathbf{u}^\mathbf{z} && (15) \\
&\geq \inf_{\boldsymbol{\alpha} \in \Delta_r : \alpha_j \geq \frac{1}{r} \text{ for some } j \notin J_\mathbf{p}} \sum_{j=1}^r \alpha_j \mathbf{p}^\top \mathbf{z}_j && (16) \\
&\geq \min_{j \notin J_\mathbf{p}} \inf_{\alpha_j \in [\frac{1}{r}, 1]} \alpha_j \mathbf{p}^\top \mathbf{z}_j + (1 - \alpha_j) \inf_{\mathbf{z} \in \mathcal{S}_\psi} \mathbf{p}^\top \mathbf{z} && (17) \\
&> \inf_{\mathbf{z} \in \mathcal{S}_\psi} \mathbf{p}^\top \mathbf{z} , && (18)
\end{aligned}$$

where the last inequality follows from Eq. (13). Since the above holds for all $\mathbf{p} \in \Delta_n$, by Lemma 2, we have that ψ is classification calibrated w.r.t. ℓ over Δ_n. $\square$

Proof of Lemma 8

Recall that a convex function $\phi : \mathbb{R}^d \to \bar{\mathbb{R}}$ (where $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$) attains its minimum at $\mathbf{u}_0 \in \mathbb{R}^d$ iff the subdifferential $\partial\phi(\mathbf{u}_0)$ contains $\mathbf{0} \in \mathbb{R}^d$ (e.g. see [18]). Also, if $\phi_1, \phi_2 : \mathbb{R}^d \to \bar{\mathbb{R}}$ are convex functions, then the subdifferential of their sum $\phi_1 + \phi_2$ at $\mathbf{u}_0$ is is equal to the Minkowski sum of the subdifferentials of ϕ_1 and ϕ_2 at $\mathbf{u}_0$:

$$\partial(\phi_1 + \phi_2)(\mathbf{u}_0) \; = \; \big\{ \mathbf{w}_1 + \mathbf{w}_2 : \mathbf{w}_1 \in \partial\phi_1(\mathbf{u}_0), \mathbf{w}_2 \in \partial\phi_2(\mathbf{u}_0) \big\} .$$

Proof. We have for all $\mathbf{p} \in \mathbb{R}^n$,

$$\begin{aligned}
\mathbf{p} \in \mathcal{N}_{\mathcal{S}_\psi}(\boldsymbol{\psi}(\hat{\mathbf{t}})) \quad &\Longleftrightarrow \quad \mathbf{p} \in \Delta_n,\ \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}) \le \mathbf{p}^\top \mathbf{z}'\ \forall \mathbf{z}' \in \mathcal{S}_\psi \\
&\Longleftrightarrow \quad \mathbf{p} \in \Delta_n,\ \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}) \le \mathbf{p}^\top \mathbf{z}'\ \forall \mathbf{z}' \in \mathcal{R}_\psi \\
&\Longleftrightarrow \quad \mathbf{p} \in \Delta_n,\ \text{and the convex function } \phi(\hat{\mathbf{t}}') = \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}') = \textstyle\sum_{y=1}^n p_y \psi_y(\hat{\mathbf{t}}') \\
&\qquad\quad \text{achieves its minimum at } \hat{\mathbf{t}}' = \hat{\mathbf{t}} \\
&\Longleftrightarrow \quad \mathbf{p} \in \Delta_n,\ \mathbf{0} \in \sum_{y=1}^n p_y \partial \psi_y(\hat{\mathbf{t}}) \\
&\Longleftrightarrow \quad \mathbf{p} \in \Delta_n,\ \mathbf{0} = \sum_{y=1}^n p_y \sum_{j=1}^{s_y} v_j^y \mathbf{w}_j^y \text{ for some } \mathbf{v}^y \in \Delta_{s_y} \\
&\Longleftrightarrow \quad \mathbf{p} \in \Delta_n,\ \mathbf{0} = \sum_{y=1}^n \sum_{j=1}^{s_y} q_j^y \mathbf{w}_j^y \text{ for some } \mathbf{q}^y = p_y \mathbf{v}^y,\ \ \mathbf{v}^y \in \Delta_{s_y} \\
&\Longleftrightarrow \quad \mathbf{p} \in \Delta_n, \mathbf{A}\mathbf{q} = \mathbf{0} \text{ for some } \mathbf{q} = (p_1 \mathbf{v}^1, \ldots, p_n \mathbf{v}^n)^\top \in \Delta_s,\ \ \mathbf{v}^y \in \Delta_{s_y} \\
&\Longleftrightarrow \quad \mathbf{p} = \mathbf{B}\mathbf{q} \text{ for some } \mathbf{q} \in \text{Null}(\mathbf{A}) \cap \Delta_s\,.
\end{aligned}$$

□

Proof of Lemma 10

Proof. For each $\hat{\mathbf{t}} \in \widehat{\mathcal{T}}$, define $\mathbf{p}^{\hat{\mathbf{t}}} = \begin{pmatrix} \hat{\mathbf{t}} \\ 1 - \sum_{j=1}^{n-1} \hat{t}_j \end{pmatrix} \in \Delta_n$. Define pred : $\widehat{\mathcal{T}} \to [k]$ as

$$\text{pred}(\hat{\mathbf{t}}) \;=\; \min\left\{ t \in [k] : \mathbf{p}^{\hat{\mathbf{t}}} \in Q_t^\ell \right\}.$$

We will show that pred satisfies the condition of Definition 1.

Fix $\mathbf{p} \in \Delta_n$. It can be seen that

$$\mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}) \;=\; \sum_{j=1}^{n-1} \Big(p_j (\hat{t}_j - 1)^2 + (1 - p_j)\, \hat{t}_j{}^2 \Big).$$

Minimizing the above over $\hat{\mathbf{t}}$ yields the unique minimizer $\hat{\mathbf{t}}^* = (p_1, \ldots, p_{n-1})^\top \in \widehat{\mathcal{T}}$, which after some calculation gives

$$\inf_{\hat{\mathbf{t}} \in \widehat{\mathcal{T}}} \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}) \;=\; \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}^*) \;=\; \sum_{j=1}^{n-1} p_j (1 - p_j)\,.$$

Now, for each $t \in [k]$, define

$$\text{regret}_{\mathbf{p}}^\ell(t) \;\triangleq\; \mathbf{p}^\top \boldsymbol{\ell}_t - \min_{t' \in [k]} \mathbf{p}^\top \boldsymbol{\ell}_{t'}\,.$$

Clearly, $\text{regret}_{\mathbf{p}}^\ell(t) = 0 \Longleftrightarrow \mathbf{p} \in \mathcal{Q}_t^\ell$. Note also that $\mathbf{p}^{\hat{\mathbf{t}}^*} = \mathbf{p}$, and therefore $\text{regret}_{\mathbf{p}}^\ell(\text{pred}(\hat{\mathbf{t}}^*)) = 0$. Let

$$\epsilon \;=\; \min_{t \in [k] : \mathbf{p} \notin \mathcal{Q}_t^\ell} \text{regret}_{\mathbf{p}}^\ell(t) \;>\; 0\,.$$

Then we have

$$\begin{aligned}
\inf_{\hat{\mathbf{t}} \in \widehat{\mathcal{T}}, \text{pred}(\hat{\mathbf{t}}) \notin \text{argmin}_t \mathbf{p}^\top \boldsymbol{\ell}_t} \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}) \quad &= \quad \inf_{\hat{\mathbf{t}} \in \widehat{\mathcal{T}}, \text{regret}_{\mathbf{p}}^\ell(\text{pred}(\hat{\mathbf{t}})) \ge \epsilon} \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}) &(19)\\
&= \quad \inf_{\hat{\mathbf{t}} \in \widehat{\mathcal{T}}, \text{regret}_{\mathbf{p}}^\ell(\text{pred}(\hat{\mathbf{t}})) \ge \text{regret}_{\mathbf{p}}^\ell(\text{pred}(\hat{\mathbf{t}}^*)) + \epsilon} \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}})\,. &(20)
\end{aligned}$$

Now, we claim that the mapping $\hat{\mathbf{t}} \mapsto \text{regret}_{\mathbf{p}}^\ell(\text{pred}(\hat{\mathbf{t}}))$ is continuous at $\hat{\mathbf{t}} = \hat{\mathbf{t}}^*$. To see this, suppose the sequence $\hat{\mathbf{t}}_m$ converges to $\hat{\mathbf{t}}^*$. Then it is easy to see that $\mathbf{p}^{\hat{\mathbf{t}}_m}$ converges to $\mathbf{p}^{\hat{\mathbf{t}}^*} = \mathbf{p}$, and therefore for each $t \in [k]$, $(\mathbf{p}^{\hat{\mathbf{t}}_m})^\top \boldsymbol{\ell}_t$ converges to $\mathbf{p}^\top \boldsymbol{\ell}_t$. Since by definition of pred we have that for all m, $\text{pred}(\hat{\mathbf{t}}_m) \in \text{argmin}_t (\mathbf{p}^{\hat{\mathbf{t}}_m})^\top \boldsymbol{\ell}_t$, this implies that for all large enough m, $\text{pred}(\hat{\mathbf{t}}_m) \in \text{argmin}_t \mathbf{p}^\top \boldsymbol{\ell}_t$. Thus for all large enough m, $\text{regret}^\ell_{\mathbf{p}}(\text{pred}(\hat{\mathbf{t}}_m)) = 0$; i.e. the sequence $\text{regret}^\ell_{\mathbf{p}}(\text{pred}(\hat{\mathbf{t}}_m))$ converges to $\text{regret}^\ell_{\mathbf{p}}(\text{pred}(\hat{\mathbf{t}}^*))$, yielding continuity at $\hat{\mathbf{t}}^*$. In particular, this implies $\exists \delta > 0$ such that

$$\|\hat{\mathbf{t}} - \hat{\mathbf{t}}^*\| < \delta \implies \text{regret}^\ell_{\mathbf{p}}(\text{pred}(\hat{\mathbf{t}})) - \text{regret}^\ell_{\mathbf{p}}(\text{pred}(\hat{\mathbf{t}}^*)) < \epsilon\,.$$

This gives

$$\inf_{\hat{\mathbf{t}} \in \widehat{\mathcal{T}}, \text{regret}^\ell_{\mathbf{p}}(\text{pred}(\hat{\mathbf{t}})) \geq \text{regret}^\ell_{\mathbf{p}}(\text{pred}(\hat{\mathbf{t}}^*)) + \epsilon} \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}) \;\geq\; \inf_{\hat{\mathbf{t}} \in \widehat{\mathcal{T}}, \|\hat{\mathbf{t}} - \hat{\mathbf{t}}^*\| \geq \delta} \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}) \tag{21}$$

$$> \inf_{\hat{\mathbf{t}} \in \widehat{\mathcal{T}}} \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}})\,, \tag{22}$$

where the last inequality holds since $\mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}})$ is a strictly convex function of $\hat{\mathbf{t}}$ and $\hat{\mathbf{t}}^*$ is its unique minimizer. The above sequence of inequalities give us that

$$\inf_{\hat{\mathbf{t}} \in \widehat{\mathcal{T}}, \text{pred}(\hat{\mathbf{t}}) \notin \text{argmin}_t \mathbf{p}^\top \boldsymbol{\ell}_t} \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}) \;>\; \inf_{\hat{\mathbf{t}} \in \widehat{\mathcal{T}}} \mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}})\,. \tag{23}$$

Since this holds for all $\mathbf{p} \in \Delta_n$, we have that ψ is classification calibrated w.r.t. ℓ over Δ_n. □

Proof of Theorem 13

The proof uses the following lemma:

Lemma 16. *Let $\ell : [n] \times [k] \to \mathbb{R}^n_+$. Let $\mathbf{p} \in \text{relint}(\Delta_n)$. Then for any $t_1, t_2 \in \arg\min_{t'} \mathbf{p}^\top \boldsymbol{\ell}_{t'}$ (i.e. such that $\mathbf{p} \in \mathcal{Q}^\ell_{t_1} \cap \mathcal{Q}^\ell_{t_2}$),*

$$\mu_{\mathcal{Q}^\ell_{t_1}}(\mathbf{p}) = \mu_{\mathcal{Q}^\ell_{t_2}}(\mathbf{p})\,.$$

Proof. Let $t_1, t_2 \in \arg\min_{t'} \mathbf{p}^\top \boldsymbol{\ell}_{t'}$ (i.e. $\mathbf{p} \in \mathcal{Q}^\ell_{t_1} \cap \mathcal{Q}^\ell_{t_2}$). Now

$$\mathcal{Q}^\ell_{t_1} = \left\{ \mathbf{q} \in \mathbb{R}^n : -\mathbf{q} \leq \mathbf{0}, \mathbf{e}^\top \mathbf{q} = 1, (\boldsymbol{\ell}_{t_1} - \boldsymbol{\ell}_t)^\top \mathbf{q} \leq 0 \;\forall t \in [k] \right\}.$$

Moreover, we have $-\mathbf{p} < \mathbf{0}$, and $(\boldsymbol{\ell}_{t_1} - \boldsymbol{\ell}_t)^\top \mathbf{p} = 0$ iff $\mathbf{p} \in \mathcal{Q}^\ell_t$. Let $\left\{ t \in [k] : \mathbf{p} \in \mathcal{Q}^\ell_t \right\} = \left\{ \tilde{t}_1, \ldots, \tilde{t}_r \right\}$ for some $r \in [k]$. Then by Lemma 14, we have

$$\mu_{Q^\ell_{t_1}} = \text{nullity}(\mathbf{A}_1)\,,$$

where $\mathbf{A}_1 \in \mathbb{R}^{(r+1) \times n}$ is a matrix containing r rows of the form $(\boldsymbol{\ell}_{t_1} - \boldsymbol{\ell}_{\tilde{t}_j})^\top, j \in [r]$ and the all ones row. Similarly, we get

$$\mu_{Q^\ell_{t_2}} = \text{nullity}(\mathbf{A}_2)\,,$$

where $\mathbf{A}_2 \in \mathbb{R}^{(r+1) \times n}$ is a matrix containing r rows of the form $(\boldsymbol{\ell}_{t_2} - \boldsymbol{\ell}_{\tilde{t}_j})^\top, j \in [r]$ and the all ones row. It can be seen that the subspaces spanned by the first r rows of $\mathbf{A}_1$ and $\mathbf{A}_2$ are both equal to the subspace parallel to the affine space containing $\boldsymbol{\ell}_{\tilde{t}_1}, \ldots, \boldsymbol{\ell}_{\tilde{t}_r}$. Thus both $\mathbf{A}_1$ and $\mathbf{A}_2$ have the same row space and hence the same null space and nullity, and therefore $\mu_{\mathcal{Q}^\ell_{t_1}}(\mathbf{p}) = \mu_{\mathcal{Q}^\ell_{t_2}}(\mathbf{p})$. □

Proof. (Proof of Theorem 13 for $\mathbf{p} \in \text{relint}(\Delta_n)$ such that $\inf_{\mathbf{z} \in \mathcal{S}_\psi} \mathbf{p}^\top \mathbf{z}$ is achieved in $\mathcal{S}_\psi$)

Let $d \in \mathbb{N}$ be such that there exists a convex surrogate target space $\widehat{\mathcal{T}} \subseteq \mathbb{R}^d$ and a convex surrogate loss $\boldsymbol{\psi} : \widehat{\mathcal{T}} \to \mathbb{R}^n_+$ that is classification calibrated with respect to ℓ over Δ_n. As noted previously, we can equivalently view $\boldsymbol{\psi}$ as being defined as $\boldsymbol{\psi} : \mathbb{R}^d \to \bar{\mathbb{R}}^n_+$, with $\psi_y(\hat{\mathbf{t}}) = \infty$ for $\hat{\mathbf{t}} \notin \widehat{\mathcal{T}}$ (and all $y \in [n]$). If $d \geq n-1$, we are done. Therefore in the following, we assume $d < n-1$.

Let $\mathbf{p} \in \text{relint}(\Delta_n)$. Note that $\inf_{\mathbf{z} \in \mathcal{S}_\psi} \mathbf{p}^\top \mathbf{z}$ always exists (since both $\mathbf{p}$ and ψ are non-negative). It can be shown that this infimum is attained in $\text{cl}(\mathcal{S}_\psi)$, i.e. $\exists \mathbf{z}^* \in \text{cl}(\mathcal{S}_\psi)$ such that $\inf_{\mathbf{z} \in \mathcal{S}_\psi} \mathbf{p}^\top \mathbf{z} = \mathbf{p}^\top \mathbf{z}^*$. In the following, we give a proof for the case when this infimum is attained within $\mathcal{S}_\psi$; the proof for the general case where the infimum is attained in $\text{cl}(\mathcal{S}_\psi)$ is similar but more technical, requiring extensions of the positive normals and the necessary condition ofTheorem 6 to sequences of points in $\mathcal{S}_\psi$ (complete details will be provided in a longer version of the paper).

For the rest of the proof, we assume $\mathbf{p}$ is such that the infimum $\inf_{\mathbf{z}\in\mathcal{S}_\psi} \mathbf{p}^\top \mathbf{z}$ is achieved in $\mathcal{S}_\psi$. In this case, it is easy to see that the infimum must then be achieved in $\mathcal{R}_\psi$ (e.g. see [18]). Thus $\exists \mathbf{z}^* = \boldsymbol{\psi}(\hat{\mathbf{t}}^*)$ for some $\hat{\mathbf{t}}^* \in \widehat{\mathcal{T}}$ such that $\inf_{\mathbf{z}\in\mathcal{S}_\psi} \mathbf{p}^\top \mathbf{z} = \mathbf{p}^\top \mathbf{z}^*$, and therefore $\mathbf{p} \in \mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z}^*)$. This gives (e.g. see discussion before proof of Lemma 8)

$$\mathbf{0} \;\in\; \partial(\mathbf{p}^\top \boldsymbol{\psi}(\hat{\mathbf{t}}^*)) = \sum_{y=1}^{n} p_y \partial\psi_y(\hat{\mathbf{t}}^*)\,.$$

Thus for each $y \in [n]$, $\exists \mathbf{w}_y \in \partial\psi_y(\hat{\mathbf{t}}^*)$ such that $\sum_{y=1}^n p_y \mathbf{w}_y = \mathbf{0}$. Now let $\mathbf{A} = \begin{bmatrix}\mathbf{w}_1 \ldots \mathbf{w}_n\end{bmatrix} \in \mathbb{R}^{d\times n}$, and let

$$\mathcal{H} \;=\; \big\{\mathbf{q} \in \Delta_n : \mathbf{A}\mathbf{q} = \mathbf{0}\big\} \;=\; \big\{\mathbf{q} \in \mathbb{R}^n : \mathbf{A}\mathbf{q} = \mathbf{0}, \mathbf{e}^\top \mathbf{q} = 1, -\mathbf{q} \le \mathbf{0}\big\}\,,$$

where $\mathbf{e}$ is the $n \times 1$ all ones vector. We have $\mathbf{p} \in \mathcal{H}$, and moreover, $-\mathbf{p} < \mathbf{0}$. Therefore, by Lemma 14, we have

$$\mu_{\mathcal{H}}(\mathbf{p}) \;=\; \text{nullity}\Big(\begin{bmatrix}\mathbf{A}\\ \mathbf{e}^\top\end{bmatrix}\Big) \;\ge\; n - (d+1)\,.$$

Now,

$$\mathbf{q} \in \mathcal{H} \implies \mathbf{A}\mathbf{q} = \mathbf{0} \implies \mathbf{0} \in \sum_{y=1}^{n} q_y \partial\psi_y(\hat{\mathbf{t}}^*) \implies \mathbf{q}^\top \mathbf{z}^* = \inf_{\mathbf{z}\in\mathcal{S}_\psi} \mathbf{q}^\top \mathbf{z} \implies \mathbf{q} \in \mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z}^*)\,,$$

which gives $\mathcal{H} \subseteq \mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z}^*)$. Moreover, by Theorem 6, we have that $\exists t_0 \in [k]$ such that $\mathcal{N}_{\mathcal{S}_\psi}(\mathbf{z}^*) \subseteq \mathcal{Q}^\ell_{t_0}$. This gives $\mathcal{H} \subseteq \mathcal{Q}^\ell_{t_0}$, and therefore

$$\mu_{\mathcal{Q}^\ell_{t_0}}(\mathbf{p}) \;\ge\; \mu_{\mathcal{H}}(\mathbf{p}) \;\ge\; n - d - 1\,.$$

By Lemma 16, we then have that for all t such that $\mathbf{p} \in \mathcal{Q}^\ell_t$,

$$\mu_{\mathcal{Q}^\ell_t}(\mathbf{p}) \;=\; \mu_{\mathcal{Q}^\ell_{t_0}}(\mathbf{p}) \;\ge\; n - d - 1\,,$$

which gives

$$d \;\ge\; n - \mu_{\mathcal{Q}^\ell_t}(\mathbf{p}) - 1\,.$$

This completes the proof for the case when $\inf_{\mathbf{z}\in\mathcal{S}_\psi} \mathbf{p}^\top \mathbf{z}$ is achieved in $\mathcal{S}_\psi$. As noted above, the proof for the case when this infimum is attained in $\text{cl}(\mathcal{S}_\psi)$ but not in $\mathcal{S}_\psi$ requires more technical details which will be provided in a longer version of the paper. ☐

Proof of Lemma 14

Proof. We will show that $\mathcal{F}_{\mathcal{C}}(\mathbf{p}) \cap (-\mathcal{F}_{\mathcal{C}}(\mathbf{p})) = \text{Null}\Big(\begin{bmatrix}\mathbf{A}^1\\ \mathbf{A}^3\end{bmatrix}\Big)$, from which the lemma follows.

First, let $\mathbf{v} \in \text{Null}\Big(\begin{bmatrix}\mathbf{A}^1\\ \mathbf{A}^3\end{bmatrix}\Big)$. Then for $\epsilon > 0$, we have

$$\begin{aligned}
\mathbf{A}^1(\mathbf{p}+\epsilon\mathbf{v}) &\;=\; \mathbf{A}^1\mathbf{p} + \epsilon\mathbf{A}^1\mathbf{v} \;=\; \mathbf{A}^1\mathbf{p} + \mathbf{0} \;=\; \mathbf{b}^1\\
\mathbf{A}^2(\mathbf{p}+\epsilon\mathbf{v}) &\;<\; \mathbf{b}^2 \quad \text{for small enough } \epsilon\text{, since } \mathbf{A}^2\mathbf{p} < \mathbf{b}^2\\
\mathbf{A}^3(\mathbf{p}+\epsilon\mathbf{v}) &\;=\; \mathbf{A}^3\mathbf{p} + \epsilon\mathbf{A}^3\mathbf{v} \;=\; \mathbf{A}^3\mathbf{p} + \mathbf{0} \;=\; \mathbf{b}^3\,.
\end{aligned}$$

Thus $\mathbf{v} \in \mathcal{F}_{\mathcal{C}}(\mathbf{p})$. Similarly, we can show $-\mathbf{v} \in \mathcal{F}_{\mathcal{C}}(\mathbf{p})$. Thus $\mathbf{v} \in \mathcal{F}_{\mathcal{C}}(\mathbf{p}) \cap (-\mathcal{F}_{\mathcal{C}}(\mathbf{p}))$, giving $\text{Null}\Big(\begin{bmatrix}\mathbf{A}^1\\ \mathbf{A}^3\end{bmatrix}\Big) \subseteq \mathcal{F}_{\mathcal{C}}(\mathbf{p}) \cap (-\mathcal{F}_{\mathcal{C}}(\mathbf{p}))$.

Now let $\mathbf{v} \in \mathcal{F}_{\mathcal{C}}(\mathbf{p}) \cap (-\mathcal{F}_{\mathcal{C}}(\mathbf{p}))$. Then for small enough $\epsilon > 0$, we have both $\mathbf{A}^1(\mathbf{p}+\epsilon\mathbf{v}) \le \mathbf{b}^1$ and $\mathbf{A}^1(\mathbf{p}-\epsilon\mathbf{v}) \le \mathbf{b}^1$. Since $\mathbf{A}^1\mathbf{p} = \mathbf{b}^1$, this gives $\mathbf{A}^1\mathbf{v} = \mathbf{0}$. Similarly, for small enough $\epsilon > 0$, we have $\mathbf{A}^3(\mathbf{p}+\epsilon\mathbf{v}) = \mathbf{b}^3$; since $\mathbf{A}^3\mathbf{p} = \mathbf{b}^3$, this gives $\mathbf{A}^3\mathbf{v} = \mathbf{0}$. Thus $\begin{bmatrix}\mathbf{A}^1\\ \mathbf{A}^3\end{bmatrix}\mathbf{v} = \mathbf{0}$, giving $\mathcal{F}_{\mathcal{C}}(\mathbf{p}) \cap (-\mathcal{F}_{\mathcal{C}}(\mathbf{p})) \subseteq \text{Null}\Big(\begin{bmatrix}\mathbf{A}^1\\ \mathbf{A}^3\end{bmatrix}\Big)$. ☐