# Appendix: Classification Calibration Dimension for General Multiclass Losses

# Calculation of Trigger Probability Sets for Figure 2

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

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

$$\begin{aligned} \mathcal{Q}_{1}^{0-1} &= \{ \mathbf{p} \in \Delta_{3} : \mathbf{p}^{\top} \boldsymbol{\ell}_{1} \leq \mathbf{p}^{\top} \boldsymbol{\ell}_{2}, \ \mathbf{p}^{\top} \boldsymbol{\ell}_{1} \leq \mathbf{p}^{\top} \boldsymbol{\ell}_{3} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{2} + p_{3} \leq p_{1} + p_{3}, \ p_{2} + p_{3} \leq p_{1} + p_{2} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{2} \leq p_{1}, \ p_{3} \leq p_{1} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{1} \geq \max(p_{2}, p_{3}) \} \end{aligned}$$

By symmetry,

$$\begin{aligned} \mathcal{Q}_{2}^{0\cdot 1} &= \{ \mathbf{p} \in \Delta_{3} : p_{2} \ge \max(p_{1}, p_{3}) \} \\ \mathcal{Q}_{3}^{0\cdot 1} &= \{ \mathbf{p} \in \Delta_{3} : p_{3} \ge \max(p_{1}, p_{2}) \} \end{aligned}$$

(b) Ordinal regression loss  $\ell^{\text{ord}}$  (n = 3).

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

$$\begin{aligned} \mathcal{Q}_{1}^{\text{ord}} &= \{ \mathbf{p} \in \Delta_{3} : \mathbf{p}^{\top} \boldsymbol{\ell}_{1} \leq \mathbf{p}^{\top} \boldsymbol{\ell}_{2}, \ \mathbf{p}^{\top} \boldsymbol{\ell}_{1} \leq \mathbf{p}^{\top} \boldsymbol{\ell}_{3} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{2} + 2p_{3} \leq p_{1} + p_{3}, \ p_{2} + 2p_{3} \leq 2p_{1} + p_{2} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{2} + p_{3} \leq p_{1}, \ p_{3} \leq p_{1} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : 1 - p_{1} \leq p_{1} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{1} \geq \frac{1}{2} \} \end{aligned}$$

By symmetry,

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

Finally,

$$\begin{aligned} \mathcal{Q}_{2}^{\text{ord}} &= \{ \mathbf{p} \in \Delta_{3} : \mathbf{p}^{\top} \boldsymbol{\ell}_{2} \leq \mathbf{p}^{\top} \boldsymbol{\ell}_{1}, \ \mathbf{p}^{\top} \boldsymbol{\ell}_{2} \leq \mathbf{p}^{\top} \boldsymbol{\ell}_{3} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{1} + p_{3} \leq p_{2} + 2p_{3}, \ p_{1} + p_{3} \leq 2p_{1} + p_{2} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{1} \leq p_{2} + p_{3}, \ p_{3} \leq p_{1} + p_{2} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{1} \leq 1 - p_{1}, \ p_{3} \leq 1 - p_{3} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{1} \leq \frac{1}{2}, \ p_{3} \leq \frac{1}{2} \} \end{aligned}$$

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

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

$$\begin{aligned} \mathcal{Q}_{1}^{(?)} &= \{ \mathbf{p} \in \Delta_{3} : \mathbf{p}^{\top} \boldsymbol{\ell}_{1} \leq \mathbf{p}^{\top} \boldsymbol{\ell}_{2}, \ \mathbf{p}^{\top} \boldsymbol{\ell}_{1} \leq \mathbf{p}^{\top} \boldsymbol{\ell}_{3}, \ \mathbf{p}^{\top} \boldsymbol{\ell}_{1} \leq \mathbf{p}^{\top} \boldsymbol{\ell}_{4} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{2} + p_{3} \leq p_{1} + p_{3}, \ p_{2} + p_{3} \leq p_{1} + p_{2}, \ p_{2} + p_{3} \leq \frac{1}{2} (p_{1} + p_{2} + p_{3}) \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{2} \leq p_{1}, \ p_{3} \leq p_{1}, \ p_{2} + p_{3} \leq \frac{1}{2} \} \\ &= \{ \mathbf{p} \in \Delta_{3} : p_{1} \geq \frac{1}{2} \} \end{aligned}$$

By symmetry,

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

Finally,

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

### **Proof of Theorem 6**

*Proof.* Since  $\psi$  is classification calibrated w.r.t.  $\ell$  over  $\Delta_n$ , by Lemma 2,  $\exists$ pred' :  $S_{\psi} \rightarrow [k]$  such that

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

Now suppose there is some  $\mathbf{z} \in S_{\psi}$  such that  $\mathcal{N}_{S_{\psi}}(\mathbf{z})$  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  $t \notin \operatorname{argmin}_{t'} \mathbf{q}^{\top} \boldsymbol{\ell}_{t'}$ . In particular, for  $t = \operatorname{pred}'(\mathbf{z}), \exists \mathbf{q} \in \mathcal{N}_{S_{\psi}}(\mathbf{z})$  such that  $\operatorname{pred}'(\mathbf{z}) \notin \operatorname{argmin}_{t'} \mathbf{q}^{\top} \boldsymbol{\ell}_{t'}$ .

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

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

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

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

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

## **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  $\mathbf{z}' \in \operatorname{conv}(\{\mathbf{z}_1, \ldots, \mathbf{z}_r\})$  and  $\mathbf{z}'' \in \mathbb{R}^n_+$ .

*Proof.* Let  $S' = \{\mathbf{z}' + \mathbf{z}'' : \mathbf{z}' \in \operatorname{conv}(\{\mathbf{z}_1, \dots, \mathbf{z}_r\}), \mathbf{z}'' \in \mathbb{R}^n_+\}$ , and suppose there exists a point  $\mathbf{z} \in S_{\psi}$  which cannot be decomposed as claimed, i.e. such that  $\mathbf{z} \notin S'$ . Then by the Hahn-Banach theorem (e.g. see [19], corollary 3.10), there exists a hyperplane that strictly separates  $\mathbf{z}$  from S', i.e.  $\exists \mathbf{w} \in \mathbb{R}^n$  such that  $\mathbf{w}^\top \mathbf{z} < \mathbf{w}^\top \mathbf{a} \ \forall \mathbf{a} \in S'$ . 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 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  $\psi$  w.r.t.  $\ell$  (over  $\Delta_n$ ) 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  $\mathbf{z} \in S_{\psi}$ ,  $\exists \boldsymbol{\alpha} \in \Delta_r, \mathbf{u} \in \mathbb{R}^n_+$  such that  $\mathbf{z} = \sum_{j=1}^r \alpha_j \mathbf{z}_j + \mathbf{u}$ . For each  $\mathbf{z} \in S_{\psi}$ , 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 pred' :  $S_{\psi} \rightarrow [k]$  as

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

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

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

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

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

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

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

Moreover, from definition of  $T_j$ , we have

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

Thus we get

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

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

$$\inf_{\mathbf{z}\in\mathcal{S}_{\psi}:\operatorname{pred}'(\mathbf{z})\notin\operatorname{argmin}_{t'}\mathbf{p}^{\top}\boldsymbol{\ell}_{t'}}\mathbf{p}^{\top}\mathbf{z} = \inf_{\mathbf{z}\in\mathcal{S}_{\psi}:\operatorname{pred}'(\mathbf{z})\notin\operatorname{argmin}_{t'}\mathbf{p}^{\top}\boldsymbol{\ell}_{t'}}\sum_{j=1}^{r}\alpha_{j}^{\mathbf{z}}\mathbf{p}^{\top}\mathbf{z}_{j} + \mathbf{p}^{\top}\mathbf{u}^{\mathbf{z}}$$
(15)

$$\geq \inf_{\boldsymbol{\alpha} \in \Delta_r: \alpha_j \geq \frac{1}{r} \text{ for some } j \notin J_{\mathbf{p}} \sum_{j=1}^{\prime} \alpha_j \mathbf{p}^\top \mathbf{z}_j$$
(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)

$$> \inf_{\mathbf{z}\in\mathcal{S}_{th}} \mathbf{p}^{\top} \mathbf{z}, \qquad (18)$$

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  $\psi$  is classification calibrated w.r.t.  $\ell$  over  $\Delta_n$ .

## **Proof of Lemma 8**

Recall that a convex function  $\phi : \mathbb{R}^d \to \overline{\mathbb{R}}$  (where  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ ) 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  $\phi_1, \phi_2 : \mathbb{R}^d \to \overline{\mathbb{R}}$  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  $\phi_1$  and  $\phi_2$  at  $\mathbf{u}_0$ :

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

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

$$\begin{split} \mathbf{p} \in \mathcal{N}_{\mathcal{S}_{\psi}}(\boldsymbol{\psi}(\hat{\mathbf{t}})) & \iff \mathbf{p} \in \Delta_{n}, \, \mathbf{p}^{\top} \boldsymbol{\psi}(\hat{\mathbf{t}}) \leq \mathbf{p}^{\top} \mathbf{z}' \, \forall \mathbf{z}' \in \mathcal{S}_{\psi} \\ & \iff \mathbf{p} \in \Delta_{n}, \, \mathbf{p}^{\top} \boldsymbol{\psi}(\hat{\mathbf{t}}) \leq \mathbf{p}^{\top} \mathbf{z}' \, \forall \mathbf{z}' \in \mathcal{R}_{\psi} \\ & \iff \mathbf{p} \in \Delta_{n}, \text{ and the convex function } \boldsymbol{\phi}(\hat{\mathbf{t}}') = \mathbf{p}^{\top} \boldsymbol{\psi}(\hat{\mathbf{t}}') = \sum_{y=1}^{n} p_{y} \psi_{y}(\hat{\mathbf{t}}') \\ & \text{achieves its minimum at } \hat{\mathbf{t}}' = \hat{\mathbf{t}} \\ & \iff \mathbf{p} \in \Delta_{n}, \, \mathbf{0} \in \sum_{y=1}^{n} p_{y} \partial \psi_{y}(\hat{\mathbf{t}}) \\ & \iff \mathbf{p} \in \Delta_{n}, \, \mathbf{0} = \sum_{y=1}^{n} p_{y} \sum_{j=1}^{s_{y}} v_{j}^{y} \mathbf{w}_{j}^{y} \text{ for some } \mathbf{v}^{y} \in \Delta_{s_{y}} \\ & \iff \mathbf{p} \in \Delta_{n}, \, \mathbf{0} = \sum_{y=1}^{n} \sum_{j=1}^{s_{y}} q_{j}^{y} \mathbf{w}_{j}^{y} \text{ for some } \mathbf{q}^{y} = p_{y} \mathbf{v}^{y}, \, \mathbf{v}^{y} \in \Delta_{s_{y}} \\ & \iff \mathbf{p} \in \Delta_{n}, \, \mathbf{A}\mathbf{q} = \mathbf{0} \text{ for some } \mathbf{q} = (p_{1}\mathbf{v}^{1}, \dots, p_{n}\mathbf{v}^{n})^{\top} \in \Delta_{s}, \, \mathbf{v}^{y} \in \Delta_{s_{y}} \\ & \iff \mathbf{p} = \mathbf{B}\mathbf{q} \text{ for some } \mathbf{q} \in \text{Null}(\mathbf{A}) \cap \Delta_{s} \,. \end{split}$$

## **Proof of Lemma 10**

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

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

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

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

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

$$\operatorname{regret}_{\mathbf{p}}^{\ell}(t) \stackrel{\Delta}{=} \mathbf{p}^{\top} \boldsymbol{\ell}_{t} - \min_{t' \in [k]} \mathbf{p}^{\top} \boldsymbol{\ell}_{t'}.$$

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

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

Then we have

$$\inf_{\hat{\mathbf{t}}\in\widehat{\mathcal{T}}, \operatorname{pred}(\hat{\mathbf{t}})\notin\operatorname{argmin}_{t}\mathbf{p}^{\top}\boldsymbol{\ell}_{t}} \mathbf{p}^{\top}\boldsymbol{\psi}(\hat{\mathbf{t}}) = \inf_{\hat{\mathbf{t}}\in\widehat{\mathcal{T}}, \operatorname{regret}_{\mathbf{p}}^{\ell}(\operatorname{pred}(\hat{\mathbf{t}})) \ge \epsilon} \mathbf{p}^{\top}\boldsymbol{\psi}(\hat{\mathbf{t}}) \qquad (19)$$

$$= \inf_{\hat{\mathbf{t}}\in\widehat{\mathcal{T}}, \operatorname{regret}_{\mathbf{p}}^{\ell}(\operatorname{pred}(\hat{\mathbf{t}})) \ge \epsilon} \mathbf{p}^{\top}\boldsymbol{\psi}(\hat{\mathbf{t}}) \dots (20)$$

$$\inf_{\hat{\mathbf{t}}\in\widehat{\mathcal{T}}, \operatorname{regret}_{\mathbf{p}}^{\ell}(\operatorname{pred}(\hat{\mathbf{t}})) \geq \operatorname{regret}_{\mathbf{p}}^{\ell}(\operatorname{pred}(\hat{\mathbf{t}}^{*})) + \epsilon} \mathbf{p}^{\top} \boldsymbol{\psi}(\hat{\mathbf{t}}) .$$
 (20)

Now, we claim that the mapping  $\hat{\mathbf{t}} \mapsto \operatorname{regret}_{\mathbf{p}}^{\ell}(\operatorname{pred}(\hat{\mathbf{t}}))$  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  $\hat{\mathbf{t}}^* = \mathbf{p}$ , 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, pred $(\hat{\mathbf{t}}_m) \in \operatorname{argmin}_t(\mathbf{p}^{\hat{\mathbf{t}}_m})^{\top} \boldsymbol{\ell}_t$ , this implies that for all large enough m, pred $(\hat{\mathbf{t}}_m) \in \operatorname{argmin}_t \mathbf{p}^{\top} \boldsymbol{\ell}_t$ . Thus for all large enough m, regret $_{\mathbf{p}}^{\ell}(\operatorname{pred}(\hat{\mathbf{t}}_m)) = 0$ ; i.e. the sequence  $\operatorname{regret}_{\mathbf{p}}^{\ell}(\operatorname{pred}(\hat{\mathbf{t}}_m))$  converges to  $\operatorname{regret}_{\mathbf{p}}^{\ell}(\operatorname{pred}(\hat{\mathbf{t}}^*))$ , yielding continuity at  $\hat{\mathbf{t}}^*$ . In particular, this implies  $\exists \delta > 0$  such that

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

This gives

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

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

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

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

### **Proof of Theorem 13**

The proof uses the following lemma:

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

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

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

$$\mu_{Q_{t_{t}}^{\ell}} = \operatorname{nullity}(\mathbf{A}_{1}),$$

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

$$\mu_{Q_{t_0}^\ell} = \operatorname{nullity}(\mathbf{A}_2),$$

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  $\mu_{\mathcal{Q}_{t_1}^{\ell}}(\mathbf{p}) = \mu_{\mathcal{Q}_{t_2}^{\ell}}(\mathbf{p})$ .  $\Box$ 

*Proof.* (Proof of Theorem 13 for  $\mathbf{p} \in \operatorname{relint}(\Delta_n)$  such that  $\inf_{\mathbf{z} \in S_{\psi}} \mathbf{p}^{\top} \mathbf{z}$  is achieved in  $S_{\psi}$ )

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  $\psi : \widehat{\mathcal{T}} \to \mathbb{R}^n_+$  that is classification calibrated with respect to  $\ell$  over  $\Delta_n$ . As noted previously, we can equivalently view  $\psi$  as being defined as  $\psi : \mathbb{R}^d \to \overline{\mathbb{R}}^n_+$ , with  $\psi_y(\hat{\mathbf{t}}) = \infty$  for  $\hat{\mathbf{t}} \notin \widehat{\mathcal{T}}$  (and all  $y \in [n]$ ). If  $d \ge n - 1$ , we are done. Therefore in the following, we assume d < n - 1.

Let  $\mathbf{p} \in \operatorname{relint}(\Delta_n)$ . Note that  $\inf_{\mathbf{z} \in S_{\psi}} \mathbf{p}^\top \mathbf{z}$  always exists (since both  $\mathbf{p}$  and  $\psi$  are non-negative). It can be shown that this infimum is attained in  $\operatorname{cl}(S_{\psi})$ , i.e.  $\exists \mathbf{z}^* \in \operatorname{cl}(S_{\psi})$  such that  $\inf_{\mathbf{z} \in S_{\psi}} \mathbf{p}^\top \mathbf{z} = \mathbf{p}^\top \mathbf{z}^*$ . In the following, we give a proof for the case when this infimum is attained within  $S_{\psi}$ ; the proof for the general case where the infimum is attained in  $\operatorname{cl}(S_{\psi})$  is similar but more technical,

requiring extensions of the positive normals and the necessary condition of Theorem 6 to sequences of points in  $S_{\psi}$  (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  $\exists \mathbf{z}^* = \psi(\hat{\mathbf{t}}^*)$  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  $\mathbf{A} = [\mathbf{w}_1 \dots \mathbf{w}_n] \in \mathbb{R}^{d \times n}$ , and let

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

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

$$\mu_{\mathcal{H}}(\mathbf{p}) = \operatorname{nullity}\left(\begin{bmatrix}\mathbf{A}\\\mathbf{e}^{\top}\end{bmatrix}\right) \geq n - (d+1).$$

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

$$\mu_{\mathcal{Q}_{t_{\alpha}}^{\ell}}(\mathbf{p}) \geq \mu_{\mathcal{H}}(\mathbf{p}) \geq n-d-1.$$

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

$$\mu_{\mathcal{Q}_t^{\ell}}(\mathbf{p}) = \mu_{\mathcal{Q}_{t_0}^{\ell}}(\mathbf{p}) \ge n - d - 1,$$

which gives

$$d \geq n - \mu_{\mathcal{Q}_{4}^{\ell}}(\mathbf{p}) - 1.$$

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

## Proof of Lemma 14

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

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

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

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  $\operatorname{Null}\left(\begin{bmatrix}\mathbf{A}^{1}\\\mathbf{A}^{3}\end{bmatrix}\right) \subseteq \mathcal{F}_{\mathcal{C}}(\mathbf{p}) \cap (-\mathcal{F}_{\mathcal{C}}(\mathbf{p}))$ .

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  $\mathbf{A}^{1}(\mathbf{p} + \epsilon \mathbf{v}) \leq \mathbf{b}^{1}$  and  $\mathbf{A}^{1}(\mathbf{p} - \epsilon \mathbf{v}) \leq \mathbf{b}^{1}$ . 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  $\mathcal{F}_{\mathcal{C}}(\mathbf{p}) \cap (-\mathcal{F}_{\mathcal{C}}(\mathbf{p})) \subseteq \operatorname{Null}(\begin{bmatrix} \mathbf{A}^{1} \\ \mathbf{A}^{3} \end{bmatrix})$ .