A Appendix

Appendix Notation

We use Δ_K to denote the subset of probability simplex i.e.,

$$\Delta_K \subset \left\{ \alpha \in \mathbb{R}^K \Big| \sum_{i=1}^{K} \alpha_i = 1, \alpha_i \geq 0 \right\}.$$

Let $\Psi_\ell(\alpha, v) = \sum_{i=1}^{K} \alpha_i \ell(v, i)$. For the ease of exposition, we define the following function: $\Psi_\ell^*(\alpha) := \inf_{v \in \Omega} \Psi_\ell(\alpha, v)$. We use $\ell_b : \Delta_K \times \mathcal{S} \to \mathbb{R}^+ \cup \{0\}$ to denote the following function: $\ell_b(\alpha, S) = \sum_{i \in [K] \setminus S} D(Y = i \mid X)$.

B Proof of Theorem 1

Proof. We generalize the result in Zhang (2004) for our proof. For the sake of clarity, we use α to denote the vector $[D(Y=1|X=x), \cdots, D(Y=K|X=x))]$. We first state few definitions and auxiliary results required for the proof. We define the following function:

$$\Delta R_{\ell_b, \Psi_\ell}(\epsilon) = \inf \left\{ \Psi_\ell(\alpha, v) - \inf_{v \in \Omega} \Psi_\ell(\alpha) \mid \ell_b(\alpha, \text{Top}_k(v)) - \inf_{v \in \Omega} \ell_b(\alpha, \text{Top}_k(v)) \geq \epsilon \right\} \cup \{+\infty\}.$$

The main idea of the proof is to show that $\Delta R_{\ell_b, \Psi_\ell}(\epsilon) > 0$ for $\epsilon > 0$. This essentially proves that the excess risk based on surrogate loss is non-zero whenever the excess Bayes risk is non-zero, also providing a bound on excess Bayes risk based on excess surrogate risk. Corollary 26 of Zhang (2004), stated below, formalizes this intuition.

Lemma 3 (Zhang (2004)). *Suppose function $\ell_b(\alpha, Top_k(v))$ is bounded and $\Delta R_{\ell_b, \Psi_\ell} > 0$ for all $\epsilon > 0$, then there exists a concave function ξ on the domain $[0, +\infty]$ that depends only on ℓ_b and Ψ_ℓ such that $\xi(0) = 0, \lim_{\epsilon \to 0^+} \xi(\epsilon) = 0$ and we have*

$$R(h) - \inf_{h' \in \mathcal{H}} R(h') \leq \xi(R_\ell(h) - \inf_{h' \in \mathcal{H}} R_\ell(h'))$$

In order to show $\Delta R_{\ell_b, \Psi_\ell}(\epsilon) > 0$ for all $\epsilon > 0$, we need the following result. This follows as a modification of Lemma 28 in Zhang (2004) and is only included here for the sake of clarity.

Lemma 4. $\forall \epsilon > 0$, $\exists \delta > 0$ *such that* $\forall \alpha \in \Delta_K$*:*

$$\inf \left\{ \Psi_\ell(\alpha, v) : v_i \leq v_{[k]} \leq v_j, \alpha_j \leq \alpha_{[k]} \leq \alpha_i, \alpha_j \leq \alpha_i - \epsilon \right\} \geq \Psi_\ell^*(\alpha) + \delta.$$

Proof. The proof is similar to Lemma 28 of Zhang (2004) except for the modification that the infimum is over the set $\{v \in \Omega \mid v_i \leq v_{[k]} \leq v_j, \alpha_j \leq \alpha_{[k]} \leq \alpha_i, \alpha_j \leq \alpha_i - \epsilon\}$. □

To prove Theorem 1, we observe the following: Suppose $\ell_b(\alpha, \text{Top}_k(v)) \geq \inf_{v \in \Omega} \ell_b(\alpha, \text{Top}_k(v)) + \epsilon$ for some $v \in \Omega$ and $\alpha \in \Delta_K$, then there $\exists i$ such that $v_i \geq v_{[k]}$ and $\alpha_i \leq \alpha_{[k]} - \epsilon$. To show this, we observe the following:

$$\ell_b(\alpha, \text{Top}_k(v)) = 1 - \sum_{j \in \text{ Top}_k(v)} \alpha_j \geq \inf_{v \in \Omega} \ell_b(\alpha, \text{Top}_k(v)) + \epsilon \geq 1 - \textstyle\sum_{j=1}^{k} \alpha_{[j]} + \epsilon,$$

and therefore, $\sum_{j \in \text{ Top}_k(v)} \alpha_j \leq \sum_{j=1}^{k} \alpha_{[j]} - \epsilon$. Note that since $|\text{Top}_k(v)| = k$, from the above inequality, it is clear that there exists $i \in \text{Top}_k(\alpha)$, $i \notin \text{Top}_k(v)$ and $j \in \text{Top}_k(v)$, $j \notin \text{Top}_k(\alpha)$ such that $\alpha_j \leq \alpha_i - \frac{\epsilon}{k}$. Furthermore, From Lemma 4, we know that $\inf\{\Psi_\ell(\alpha, v) : v_i \leq v_{[k]} \leq v_j, \alpha_i \geq \alpha_{[k]} \geq \alpha_j, \alpha_j \leq \alpha_i - \frac{\epsilon}{k}\} \geq \Psi_\ell^*(\alpha) + \delta$. Therefore, $\Delta R_{\ell_b, \Psi_\ell}(\epsilon) > 0$. Using Lemma 3, we get the required result. □

C Proof of Lemma 2

Proof. The fact that ℓ_{snm} is a POWL or BOWL is evident from the formula for the random variable $L(v, y)$.

Let σ be a permutation of $[K-1]$ which sorts the coordinates of v^{-y} in non-increasing order, i.e. $v^{-y}_{\sigma(j)} \geq v^{-y}_{\sigma(j')}$ for $j < j'$. Then we have Note that

$$\begin{aligned}
\theta_j &= \mathbb{E}_{\mathcal{B}}\left[\sum_{i=1}^{B} \vartheta_i \mathbb{I}(\sigma(j) \in \mathcal{B} \text{ and } v_{\sigma(j)} \text{ is the } i^{\text{th}} \text{ largest score in } \mathcal{B})\right] \\
&= \sum_{i=1}^{B} \vartheta_i \Pr_{\mathcal{B}}[\sigma(j) \in \mathcal{B} \text{ and } v_{\sigma(j)} \text{ is the } i^{\text{th}} \text{ largest score in } \mathcal{B}] \\
&= \sum_{i=1}^{B} \vartheta_i \Pr_{\mathcal{B}}[v_{\sigma(j)} \text{ is the } i^{\text{th}} \text{ largest score in } \mathcal{B} | \sigma(j) \in \mathcal{B}] \cdot \frac{B}{K-1}. \qquad (4)
\end{aligned}$$

Now, if $j' > j$, then since $v_{\sigma(j')} \leq v_{\sigma(j)}$, we have

$$Pr_{\mathcal{B}}[v_{\sigma(j)} \text{ is the } i^{\text{th}} \text{ largest score in } \mathcal{B}|\sigma(j) \in \mathcal{B}] \geq Pr_{\mathcal{B}}[v_{\sigma(j')} \text{ is the } i^{\text{th}} \text{ largest score in } \mathcal{B}|\sigma(j') \in \mathcal{B}].$$

This is easy to check by comparing the two events. Since the coordinates of ϑ are non-increasing, this implies that $\theta_j \geq \theta_{j'}$, thus establishing that the coordinates of θ are also non-increasing.

Next, suppose that $\vartheta_i = \frac{K-1}{kB}$ for $i \in [k]$. Let $j \in [k]$. Note that if $\sigma(j) \in \mathcal{B}$, then $v_{\sigma(j)}$ is among the top k scores in $\mathcal{B}$. Thus by (4), we conclude that $\theta_j = {}^1\!/_k$.

Finally, if $\vartheta_i > 0$ for all $i \in [B]$, then by (4), we have $\theta_j > 0$.

□

D Proofs of Theorems 2 and 3

Proof of Theorem 2. Consider the (ϕ, θ)-POWL ℓ. Fix any class $y \in \mathcal{Y}$. Since ϕ is a non-increasing function, we have $\phi(v_y - v^{-y}_{[j]}) \geq \phi(v_y - v^{-y}_{[j']})$ if $j < j'$. Since θ has non-increasing coordinates, by the Rearrangement Inequality, we conclude that for any permutation σ of $[K-1]$, we have

$$\sum_{j=1}^{K-1} \theta_{\sigma(j)} \phi(v_y - v^{-y}_j) \leq \sum_{j=1}^{K-1} \theta_j \phi(v_y - v^{-y}_{[j]}) = \ell(v, y).$$

Since the above inequality holds for any permutation σ, we have

$$\ell(v, y) = \max_{\sigma} \sum_{j=1}^{K-1} \theta_{\sigma(j)} \phi(v_y - v^{-y}_j).$$

Note that $v \mapsto \sum_{j=1}^{K-1} \theta_{\sigma(j)} \phi(v_y - v^{-y}_j)$ is a convex function of v since it is non-negative linear combination of convex functions of v. Hence $\ell(v, y)$ is a convex function of v since it is the maximum of convex functions of v.

The proof that the (ϕ, θ)-BOWL is also convex is very similar and is omitted for brevity. □

Proof Theorem 3. First, consider the (ϕ, θ)-POWL ℓ. Suppose $y \notin \text{Top}_k(v)$. Then for any $j \in [k]$, we have $v_y \leq v^{-y}_{[j]}$, and so $\phi(v_y - v^{-y}_{[j]}) \geq \mathbb{I}(v_y - v^{-y}_{[j]} \leq 0) = 1$. Since θ is a non-negative vector and ϕ is also non-negative, we have

$$\ell(v, y) \geq \sum_{j=1}^{k} \theta_j \phi(v_y - v^{-y}_{[j]}) \geq \sum_{j=1}^{k} \theta_j \cdot 1 = \mathbb{I}(y \notin \text{Top}_k(v)).$$

If $y \in \text{Top}_k(v)$, then $\mathbb{I}(v \notin \text{Top}_k(v)) = 0$, and $\ell(v, y) \geq \mathbb{I}(v \notin \text{Top}_k(v))$ since $\ell(v, y)$ is always non-negative.

Now, consider the (ϕ, θ)-BOWL ℓ. We have

$$\ell(v, y) = \phi(v_y) + \sum_{j=1}^{K-1} \theta_j \phi(v_y - v_{[j]}^{-y}) \geq \sum_{j=1}^{k} \theta_j (\phi(v_y) + \phi(-v_{[j]}^{-y})) \geq \sum_{j=1}^{k} 2\theta_j (\phi(\tfrac{1}{2}(v_y - v_{[j]}^{-y}))).$$

The first inequality above follows since $\theta_j = 1/k$ for $j \in [k]$ and the fact that ϕ is always non-negative, and the second inequality by the convexity of ϕ. Now arguing just like in the POWL case, we have

$$\sum_{j=1}^{k} 2\theta_j (\phi(\tfrac{1}{2}(v_y - v_{[j]}^{-y}))) \geq 2\mathbb{I}(v \notin \mathrm{Top}_k(v)).$$

□

E Proof of Theorem 4

Proof. We first prove the following key order-preserving property of the loss functions in Definition 3 and 4 (the proof of the result is given in Lemma 5 and Lemma 6).

Lemma. *Suppose ϕ satisfies the conditions in Theorem 4. Then for any $\alpha \in \Delta_K$ that satisfies the following condition:*

$$\alpha_{[k]} > \frac{\sum_{l=k+1}^{k+q} \alpha_{[l]}}{k \sum_{j=k}^{k+q-1} \theta_j},$$

for all $q \in [K-k]$ and $v \in \mathbb{R}^K$ such that $\Psi_\ell(\alpha, v) = \Psi_\ell^*(\alpha)$ for ℓ in Definition 3 and Definition 4 with appropriate conditions on $\{\theta_i\}_{i=1}^{K-1}$ (as specified in Theorem 4), we have

1. *$v_i \geq v_j$ when $\alpha_i > \alpha_j$ and*
2. *$v_{[i]} > v_{[j]}$ when $\alpha_i > \alpha_j$ and $i \in [k]$ and $j \in [K] \backslash [k]$.*

The proof can be completed by appealing to the order preserving property of Ψ_ℓ in the above lemma. In particular, consider v' such that $\Psi_\ell(\alpha, v') = \Psi_\ell^*(\alpha)$, then it is shown that $v_{[i]} > v_{[j]}$ when $\alpha_i > \alpha_j$ and $i \in [k]$ and $j \in [K] \backslash [k]$. From this result, it is easy to see that $\lim_{t \to \infty} \Psi_\ell(\alpha, v^t) = \Psi_\ell(\alpha, v) > \Psi_\ell(\alpha, v') = \inf_{v \in \Omega} \Psi_\ell(\alpha, v) = \Psi_\ell^*(\alpha)$, thus, completing the proof. □

E.1 Lemmatta for Theorem 4

Lemma 5. *Suppose ϕ satisfies the conditions in Theorem 4. Then for any $\alpha \in \Delta_K$ that satisfies the following condition:*

$$\alpha_{[k]} > \frac{\sum_{l=k+1}^{k+q} \alpha_{[l]}}{k \sum_{j=k+1}^{k+q} \theta_j},$$

for all $q \in [K-k]$ and $v \in \mathbb{R}^K$ such that $\Psi_\ell(\alpha, v) = \Psi_\ell^*(\alpha)$ for ℓ in Definition 3 with $\theta_j = \frac{1}{k}$ for all $j \in [k]$ and $\theta_j \leq \frac{1}{k}$ for $j > k$, we have

1. *$v_i \geq v_j$ when $\alpha_i > \alpha_j$ and*
2. *$v_{[i]} > v_{[j]}$ when $\alpha_i > \alpha_j$ and $i \in [k]$ and $j \in [K] \backslash [k]$.*

Proof. We prove the first part by contradiction. Assume $\exists j_1, j_2$ such that $\alpha_{j_1} > \alpha_{j_2}$ but $v_{j_1} < v_{j_2}$. Consider $\bar{v}$

such that $\bar{v}_i = v_i$ for all $i \neq j_1, j_2$, $\bar{v}_{j_1} = v_{j_2}$ and $\bar{v}_{j_2} = v_{j_1}$. Then we have

$$\begin{aligned}
&\Psi_\ell(\alpha, \bar{v}) - \Psi_\ell(\alpha, v) \\
&= \alpha_{j_1} \left(\sum_{j=1}^{K-1} \theta_j \phi(\bar{v}_{j_1} - \bar{v}_{[j]}^{-j_1}) - \sum_{j=1}^{K-1} \theta_j \phi(v_{j_1} - v_{[j]}^{-j_1}) \right) + \alpha_{j_2} \left(\sum_{j=1}^{K-1} \theta_j \phi(\bar{v}_{j_2} - \bar{v}_{[j]}^{-j_2}) - \sum_{j=1}^{K-1} \theta_j \phi(v_{j_2} - v_{[j]}^{-j_2}) \right) \\
&= (\alpha_{j_1} - \alpha_{j_2}) \left(\sum_{j=1}^{K-1} \theta_j \phi(v_{j_2} - v_{[j]}^{-j_2}) - \sum_{j=1}^{K-1} \theta_j \phi(v_{j_1} - v_{[j]}^{-j_1}) \right)
\end{aligned}$$

The above equality is due to the definition of $\bar{v}$. Furthermore, we observe the following: $v_{j_2} > v_{j_1}$ and $v_{[j]}^{-j_1} \geq v_{[j]}^{-j_2}$ for all $j \in [K-1]$. This is due to the fact that removal of v_{j_2} rather than v_{j_1} from v can only decrease the order statistic.. Therefore, we have

$$v_{j_2} - v_{[j]}^{-j_2} > v_{j_1} - v_{[j]}^{-j_1},$$

for all $j \in [K-1]$. Since ϕ is non-increasing, it is clear that $\Psi_\ell(\alpha, \bar{v}) - \Psi_\ell(\alpha, v) \leq 0$. Also, note that at least one $v_{j_1} - v_{[j]}^{-j_1} < 0$ since $v_{j_2} > v_{j_1}$ for $j \in [k]$. Since ϕ is strictly decreasing on $(-\infty, 0]$, we can, in fact, obtain $\Psi_\ell(\alpha, \bar{v}) - \Psi_\ell(\alpha, v) < 0$, which is a contradiction to the optimality of v.

We now focus on the second part of the proof. Without loss of generality, suppose $\alpha_1 \geq \cdots \geq \alpha_k > \alpha_{k+1} \geq \cdots \alpha_K$. Suppose $v_k > v_{k+1}$, then the second part follows immediately. Now, consider the scenario:

$$v_1 \geq v_2 \geq \cdots \geq v_k = v_{k+1} = \cdots = v_{k+q} > v_{k+q+1} \geq \cdots \geq v_K.$$

We will prove that such a scenario is not possible. We prove this by contradiction. Consider the vector v' defined as follows:

$$v'_i = \begin{cases} v_i + \delta, & \text{for } i = k \\ v_i - \beta\delta, & \text{for } k+1 \leq i \leq k+q \\ v_i, & \text{otherwise .} \end{cases}$$

Here δ is chosen sufficiently small such that $v'_{k+q} > v'_{k+q+1}$ with $\beta = \frac{1}{k \sum_{j=k+1}^{k+q} \theta_j}$. When α, v are held fixed, with slight abuse of notation, we use $\Psi_\ell(\delta)$ to denote part of the function $\Psi_\ell(\alpha, v')$ that only depends on δ. Let us denote the remaning part by $C_{\alpha,v}$ such that $\Psi_\ell(\alpha, v) = \Psi_\ell(0) + C_{\alpha,v}$. More specifically, we have the following:

$$\begin{aligned}
\Psi_\ell(\delta) = &\underbrace{\alpha_k \left[\sum_{j=1}^{k-1} \frac{1}{k} \phi(v_k - v_j + \delta) + \sum_{j=k+q+1}^{K} \theta_{j-1} \phi(v_k - v_j + \delta) + \sum_{j=k+1}^{k+q} \theta_{j-1} \phi(v_k - v_j + (1+\beta)\delta) \right]}_{T_1(\delta)} \\
&+ \underbrace{\sum_{l=k+1}^{k+q} \alpha_l \left[\sum_{j=1}^{k-1} \frac{1}{k} \phi(v_l - v_j - \beta\delta) + \sum_{j=k+q+1}^{K} \theta_{j-1} \phi(v_l - v_j - \beta\delta) + \frac{1}{k} \phi(v_l - v_k - (1+\beta)\delta) \right]}_{T_2(\delta)} \\
&+ \underbrace{\sum_{l=1}^{k-1} \alpha_l \left[\frac{1}{k} \phi(v_l - v_k - \delta) + \sum_{j=k+1}^{k+q} \theta_{j-1} \phi(v_l - v_j + \beta\delta) \right]}_{T_3(\delta)} \\
&+ \underbrace{\sum_{l=k+q+1}^{K} \alpha_l \left[\frac{1}{k} \phi(v_l - v_k - \delta) + \sum_{j=k+1}^{k+q} \theta_j \phi(v_l - v_j + \beta\delta) \right]}_{T_4(\delta)}
\end{aligned}$$

Since, $\theta_i = 1/k$ for all $i \leq k$, $\Psi_\ell(\alpha, v') = \Psi_\ell(\delta) + C_{\alpha,v}$ for $0 \leq \beta\delta \leq v'_{k+q} - v'_{k+q+1}$. This follows the fact the the rank (position when sorted) of of v'_i amongst elements in v' is same as that of v_i amongst elements in v for $i > k$ for sufficiently small chosen δ since the rank of v'_k in v' can only decrease in comparison to rank v_k in v and the rank remains same for all $i > k$. Also, note that Ψ_ℓ is differentiable. Our aim is to show that $\Psi'_\ell(0) < 0$, which implies a contradiction to the optimality of v. To this end, we analyze the differential of aforementioned terms separately as follows:

$$\begin{aligned} T'_3(\delta) &= \sum_{l=1}^{k-1} \alpha_l \left[-\frac{1}{k}\phi'(v_l - v_k - \delta) + \beta \sum_{j=k+1}^{k+q} \theta_{j-1}\phi'(v_l - v_j + \beta\delta) \right] \\ &= \sum_{l=1}^{k-1} \alpha_l \left[-\frac{1}{k}\phi'(v_l - v_k - \delta) + \phi'(v_l - v_k + \beta\delta)\beta \sum_{j=k+1}^{k+q} \theta_{j-1} \right]. \end{aligned}$$

The above equality holds because $v_k = v_i$ for all $i \in [k+1, k+q]$. From the above equality, we have:

$$T'_3(0) = \sum_{l=1}^{k-1} \alpha_l \left[\left(\beta \sum_{j=k+1}^{k+q} \theta_{j-1} - \frac{1}{k} \right) \phi'(v_l - v_k) \right] \leq 0.$$

This is due to the fact that ϕ is non-increasing and following inequality

$$\beta \sum_{j=k+1}^{k+q} \theta_{j-1} \geq \frac{1}{k}.$$

In a similar manner, it can also be shown that $T'_4(0) = 0$. To complete the proof, we need to show that $T'_1(0) + T'_2(0) < 0$. We observe the following:

$$\begin{aligned} T'_1(\delta) + T'_2(\delta) &= \frac{1}{k}\sum_{j=1}^{k-1} \left(\alpha_k \phi'(v_k - v_j + \delta) - \phi'(v_k - v_j - \beta\delta)\beta \sum_{l=k+1}^{k+q} \alpha_l \right) \\ &\quad + \sum_{j=k+q+1}^{K} \theta_{j-1} \left(\alpha_k \phi'(v_k - v_j + \delta) - \phi'(v_k - v_j - \beta\delta)\beta \sum_{l=k+1}^{k+q} \alpha_l \right) \\ &\quad + \phi'((1+\beta)\delta)(1+\beta)\alpha_k \sum_{j=k+1}^{k+q} \theta_{j-1} - \frac{1+\beta}{k}\phi'(-(1+\beta)\delta) \sum_{l=k+1}^{k+q} \alpha_l \end{aligned}$$

The above equality is due to the fact that $v_k = v_i$ for all $i \in [k+1, k+q]$. From the above equality we have,

$$\begin{aligned} T'_1(0) + T'_2(0) &= \frac{1}{k}\sum_{j=1}^{k-1} \phi'(v_k - v_j) \left(\alpha_k - \beta \sum_{l=k+1}^{k+q} \alpha_l \right) \\ &\quad + \sum_{j=k+q+1}^{K} \phi'(v_k - v_j)\theta_{j-1} \left(\alpha_k - \beta \sum_{l=k+1}^{k+q} \alpha_l \right) \\ &\quad + (1+\beta)\phi'(0) \left[\alpha_k \sum_{j=k+1}^{k+q} \theta_{j-1} - \frac{1}{k} \sum_{l=k+1}^{k+q} \alpha_l \right] \end{aligned}$$

From the above equality, we can see that $T'_1(0) + T'_2(0) < 0$. This is due to the fact that ϕ is non-increasing with $\phi'(0) < 0$ and the following inequalities:

$$\begin{aligned} &\alpha_k > \beta \sum_{l=k+1}^{k+q} \alpha_l = \frac{\sum_{l=k+1}^{k+q} \alpha_l}{k \sum_{l=k+1}^{k+q} \theta_j} \\ &\alpha_k \sum_{j=k}^{k+q-1} \theta_j > \frac{1}{k} \sum_{l=k+1}^{k+q} \alpha_l. \end{aligned}$$

Therefore, we have $\Psi'_\ell(0) = T'_1(0) + T'_2(0) + T'_3(0) + T'_4(0) < 0$. This is a contradiction to the optimality of v. Hence, the scenario

$$v_1 \geq v_2 \geq \cdots \geq v_k = v_{k+1} = \cdots = v_{k+q} > v_{k+q+1} \geq \cdots \geq v_K,$$

is not possible. This completes the proof of second part of the lemma. □

Lemma 6. *Suppose ϕ satisfies the conditions in Theorem 4. Then for any $\alpha \in \Delta_K$ that satisfies the following condition:*

$$\alpha_{[k]} > \frac{\sum_{l=k+1}^{k+q} \alpha_{[l]}}{k \sum_{j=k+1}^{k+q} \theta_j},$$

for all $q \in [K-k]$ and $v \in \mathbb{R}^K$ such that $\Psi_\ell(\alpha, v) = \Psi^*_\ell(\alpha)$ for ℓ in Definition 4 with $\theta_j = \frac{1}{k}$ for all $j \in [k]$ and $\theta_j \leq \frac{1}{k}$ for $j > k$, we have

1. *$v_i \geq v_j$ when $\alpha_i > \alpha_j$ and*

2. *$v_{[i]} > v_{[j]}$ when $\alpha_i > \alpha_j$ and $i \in [k]$ and $j \in [K] \backslash [k]$.*

Proof. We prove the first part by contradiction. Assume $\exists j_1, j_2$ such that $\alpha_{j_1} > \alpha_{j_2}$ but $v_{j_1} < v_{j_2}$. Consider $\bar{v}$ such that $\bar{v}_i = v_i$ for all $i \neq j_1, j_2$, $\bar{v}_{j_1} = v_{j_2}$ and $\bar{v}_{j_2} = v_{j_1}$. Then we have

$$\Psi_\ell(\alpha, \bar{v}) - \Psi_\ell(\alpha, v) = (\alpha_{j_1} - \alpha_{j_2}) \left(\phi(v_{j_2}) + \sum_{j=1}^{K-1} \theta_j \phi(-v_{[j]}^{-j_2}) - \phi(v_{j_1}) - \sum_{j=1}^{K-1} \theta_j \phi(-v_{[j]}^{-j_1}) \right)$$

The above equality is due to the definition of $\bar{v}$. Furthermore, we observe the following: $v_{j_2} > v_{j_1}$ and $v_{[j]}^{-j_1} \geq v_{[j]}^{-j_2}$ for all $j \in [K-1]$. This is due to the fact that removal of v_{j_2} rather than v_{j_1} from v can only decrease the order statistic. If v_{j_1} is non-positive, then $\phi(v_{j_2}) < \phi(v_{j_1})$ and $\phi(-v_{[j]}^{-j_2}) \leq \phi(-v_{[j]}^{-j_1})$ as $\phi'(\epsilon) < 0$ for all $\epsilon \leq 0$ and ϕ is non-increasing, which is a contradiction to the optimality of v.

We now consider the case where $v_{j_2} > v_{j_1} > 0$. It is not hard to see that $v_{[j]}^{-j_1} = v_{[j]}^{-j_2}$ whenever $v_{[j]}^{-j_1} < v_{j_1}$. Furthermore, $\sum_{i=1}^{K-1} v_{[j]}^{-j_1} > \sum_{i=1}^{K-1} v_{[j]}^{-j_2}$. From the above two facts, we get $v_{[j]}^{-j_1} > v_{[j]}^{-j_2}$ for some j such that $v_{[j]}^{-j_2} > 0$. For this j, $\phi(-v_{[j]}^{-j_2}) \leq \phi(-v_{[j]}^{-j_1})$ as $\phi'(\epsilon) < 0$ for all $\epsilon \leq 0$. Since ϕ is strictly decreasing on $(-\infty, 0]$, we can, in fact, obtain $\Psi_\ell(\alpha, \bar{v}) - \Psi_\ell(\alpha, v) < 0$, which is again a contradiction to the optimality of v. This completes the first part of the proof.

We now turn our attention to the second part. For the ease of exposition, suppose $\alpha_1 \geq \cdots \geq \alpha_k > \alpha_{k+1} \geq \cdots \alpha_K$. The proof is along similar lines as that of pairwise comparison method. Suppose $v_k > v_{k+1}$, then the second part follows immediately. Now, consider the scenario:

$$v_1 \geq v_2 \geq \cdots \geq v_k = v_{k+1} = \cdots = v_{k+q} > v_{k+q+1} \geq \cdots \geq v_K.$$

We will prove that is not possible through proof by contradiction. Consider the vector v' defined as follows:

$$v'_i = \begin{cases} v_i + \delta, & \text{for } i = k \\ v_i - \beta\delta, & \text{for } k+1 \leq i \leq k+q \\ v_i, & \text{otherwise }, \end{cases}$$

where δ is chosen sufficiently small such that $v'_{k+q} > v'_{k+q+1}$ with $\beta = \frac{1}{k \sum_{j=k+1}^{k+q} \theta_j}$. When α, v are held fixed, with slight abuse of notation, we use $\Psi_\ell(\delta)$ to denote part of the function $\Psi_\ell(\alpha, v')$ that only depends on δ. Let us denote the remaning part by $C_{\alpha,v}$ such that $\Psi_\ell(\alpha, v) = \Psi_\ell(0) + C_{\alpha,v}$. More specifically, we have the following:

$$\Psi_\ell(\delta) = \underbrace{\alpha_k \left[\phi(v_k+\delta) + \sum_{j=k+1}^{k+q} \theta_{j-1}\phi(-v_j+\beta\delta)\right]}_{T_1(\delta)}$$
$$+ \underbrace{\sum_{l=k+1}^{k+q} \alpha_l \left[\phi(v_l-\beta\delta) + \frac{1}{k}\phi(-v_k-\delta) + \sum_{j=k+1}^{l-1}\theta_j\phi(-v_j+\beta\delta) + \sum_{j=l+1}^{k+q}\theta_{j-1}\phi(-v_j+\beta\delta)\right]}_{T_2(\delta)}$$
$$+ \underbrace{\sum_{l=1}^{k-1}\alpha_l\left[\frac{1}{k}\phi(-v_k-\delta) + \sum_{j=k+1}^{k+q}\theta_{j-1}\phi(-v_j+\beta\delta)\right]}_{T_3(\delta)} + \underbrace{\sum_{l=k+q+1}^{K}\alpha_l\left[\frac{1}{k}\phi(-v_k-\delta) + \sum_{j=k+1}^{k+q}\theta_j\phi(-v_j+\beta\delta)\right]}_{T_4(\delta)}$$

Since, $\theta_i = \frac{1}{k}$ for all $i \le k$, $\Psi_\ell(\alpha, v') = \Psi_\ell(\delta) + C_{\alpha,v}$ for $0 \le \beta\delta \le v'_{k+q} - v'_{k+q+1}$ and $\Psi_\ell(\delta)$ is differentiable as argued for POWL. Our goal is to show that $\Psi'_\ell(0) < 0$, which implies $\Psi_\ell(\alpha, v') < \Psi_\ell(\alpha, v)$, thereby contradicting the optimality of v. With our choice of β, it can be shown that $T'_3(0) \le 0$ and $T'_4(0) = 0$ using the same argument for corresponding terms for POWL. To complete the proof, we need to show that $T'_1(0) + T'_2(0) < 0$. We observe the following:

$$T'_1(\delta) + T'_2(\delta) = \alpha_k\phi'(v_k+\delta) + \beta\alpha_k\sum_{j=k+1}^{k+q}\theta_{j-1}\phi'(-v_k+\beta\delta)$$
$$+ \sum_{l=k+1}^{k+q}\alpha_l\left[-\beta\phi'(v_k-\beta\delta) - \frac{1}{k}\phi'(-v_k-\delta) + \beta\sum_{j=k+1}^{k+q-1}\theta_j\phi'(-v_k+\beta\delta)\right]$$

The above equality is due to the fact that $v_k = v_i$ for all $i \in [k+1, k+q]$. From the above equality we have,

$$T'_1(0) + T'_2(0) = \alpha_k\phi'(v_k) + \beta\alpha_k\sum_{j=k+1}^{k+q}\theta_{j-1}\phi'(-v_k) + \sum_{l=k+1}^{k+q}\alpha_l\left[-\beta\phi'(v_k) - \frac{1}{k}\phi'(-v_k) + \beta\sum_{j=k+1}^{k+q-1}\theta_j\phi'(-v_k)\right]$$
$$= \left(\alpha_k - \beta\sum_{l=k+1}^{k+q}\alpha_l\right)\phi'(v_k) + \sum_{l=k+1}^{k+q}\left(\alpha_k\sum_{l=k+1}^{k+q}\alpha_l - \theta_{k+q}\sum_{l=k+1}^{k+q}\alpha_\ell\right)\beta\phi'(-v_k) < 0.$$

The last inequality is due to the following:

$$\alpha_k > \beta\sum_{l=k+1}^{k+q}\alpha_l = \frac{\sum_{l=k+1}^{k+q}\alpha_l}{k\sum_{l=k+1}^{k+q}\theta_j}.$$

$\theta_{k+q} \le \frac{1}{k}$ and the fact that at least one of $\phi'(-v_k)$ and $\phi'(v_k)$ is strictly negative as $\phi'(\epsilon) < 0$ for $\epsilon \le 0$. Therefore, we have $\Psi'_\ell(0) = T'_1(0) + T'_2(0) + T'_3(0) + T'_4(0) < 0$. This is a contradiction to the optimality of v. Hence, the scenario

$$v_1 \ge v_2 \ge \cdots \ge v_k = v_{k+1} = \cdots = v_{k+q} > v_{k+q+1} \ge \cdots \ge v_K,$$

is not possible. This completes the proof of second part of the lemma. □

F Proofs of Theorems 5 and 6

Proof of Theorem 5. Our generalization bounds are based on the work of Lei et al. (2015), who give general purpose bounds in terms of Lipschitz constants and range of the loss. In particular, suppose that $|\ell(v, y)| \le \Phi'$. Further, suppose that for any $y \in \mathcal{Y}$, ℓ satisfies an L_2-Lipschitzness condition of the form:

$$|\ell(v, y) - \ell(u, y)| \le L_1\|v - u\|_2 + L_2|v_y - u_y|,$$

and an L_∞-Lipschitzness condition of the form:

$$|\ell(v,y) - \ell(u,y)| \leq L_3 \|v - u\|_\infty.$$

Then Lei et al. (2015) prove (see Theorems 2 and 6 in their paper[5]) that the generalization error is bounded with probability at least $1-\delta$ by

$$\tilde{O}\left(\min\left\{L_1 K \mathfrak{G}_{\bar{S}}(\mathcal{H}) + L_2 \mathfrak{G}_S(\mathcal{H}), L_3\sqrt{K}\mathfrak{R}_{\bar{S}}(\mathcal{H})\right\}\right) + 3\Phi'\sqrt{\frac{\log(2/\delta)}{2n}}.$$

For OWLs, Lemma 7 provides the required Lipschitz constants. Next, it is easy to check that the setting $\Phi' = \|\theta\|_1 \Phi$ is a valid bound on the range of the losses. The claimed generalization bound follows by plugging in the values of the Lipschitz constants and Φ'. □

Lemma 7. *Let $\phi(\cdot)$ be L-Lipschitz. Let $u, v \in \mathbb{R}^K$ be two score vectors. Then the (ϕ,θ)-POWL ℓ satisfies the following Lipschitzness conditions, for any $y \in \mathcal{Y}$:*

$$|\ell(v,y) - \ell(u,y)| \leq \begin{cases} L\|\theta\|_2\|v-u\|_2 + L\|\theta\|_1|v_y - u_y| & (L_2\text{-Lipschitzness}) \\ 2L\|\theta\|_1\|v-u\|_\infty & (L_\infty\text{-Lipschitzness}) \end{cases}$$

Furthermore, the (ϕ,θ)-BOWL ℓ satisfies the following Lipschitzness conditions, for any $y \in \mathcal{Y}$:

$$|\ell(v,y) - \ell(u,y)| \leq \begin{cases} L\|\theta\|_2\|v-u\|_2 + L|v_y - u_y| & (L_2\text{-Lipschitzness}) \\ L(\|\theta\|_1 + 1)\|v-u\|_\infty & (L_\infty\text{-Lipschitzness}) \end{cases}$$

Proof. We first consider the (ϕ,θ)-POWL ℓ. Let $p \in \{2, \infty\}$. Then we have

$$\begin{aligned} |\ell(v,y) - \ell(u,y)| &= \left|\sum_{j=1}^{K-1} \theta_j(\phi(v_y - v^{-y}_{[j]}) - \phi(u_y - u^{-y}_{[j]}))\right| \\ &\leq \sum_{j=1}^{K-1} \theta_j \cdot L(|v_y - u_y| + |v^{-y}_{[j]} - u^{-y}_{[j]}|) \\ &\leq L\|\theta\|_1|v_y - u_y| + L\|\theta\|_p\|\tilde{v}^{-y} - \tilde{u}^{-y}\|_{p/(p-1)}. \end{aligned}$$

The first inequality above follows from the L-Lipschitzness of ϕ and the triangle inequality, and the second by Hölder's inequality. Then applying the bounds from Lemma 8, we get the claimed bounds.

The claimed bounds for the (ϕ,θ)-BOWL are obtained using an almost identical analysis and is omitted for brevity. □

Lemma 8. *Let $u, v \in \mathbb{R}^K$ be two score vectors, and let $\tilde{u}, \tilde{v} \in \mathbb{R}^K$ be sorted versions of u, v respectively with coordinates in non-increasing order. Then we have*

$$\|\tilde{v} - \tilde{u}\|_2 \leq \|v - u\|_2 \quad \text{and} \quad \|\tilde{v} - \tilde{u}\|_\infty \leq \|v - u\|_\infty.$$

Proof. The first inequality is an easy consequence of the Rearrangement Inequality after squaring both sides. As for the second inequality, let $\epsilon := \|v - u\|_\infty$, and let $k \in \mathcal{Y}$ be any index. Then note that for any $j \in \text{Top}_k(u)$, we have $v_j \geq u_j - \epsilon$, and hence $\tilde{v}_k \geq \tilde{u}_k - \epsilon$. Similarly, $\tilde{u}_k \geq \tilde{v}_k - \epsilon$. These two inequalities imply that $|\tilde{v}_k - \tilde{u}_k| \leq \epsilon$, and thus the claimed bound follows. □

Proof of Theorem 6. Consider the $(\phi_{\text{ramp},\rho}, \theta)$-POWL ℓ where $\theta_k = 1$ and $\theta_j = 0$ for all $j \neq k$. Thus, this loss can be rewritten as $\ell(v,y) = \phi_{\text{ramp},\rho}(v_y - v^{-y}_{[k]})$, and hence for a given hypothesis h and an example (x,y), we have $\ell(h(x), y) = \phi_{\text{ramp},\rho}(\rho_h(x,y))$. The claimed margin bound then follows by applying the bound from Theorem 5 using the facts that $\|\theta\|_1 = \|\theta\|_2 = 1$, $L = \frac{1}{\rho}$, $\Phi = 1$, and $\mathbb{I}[u \leq 0] \leq \phi_{\text{ramp},\rho}(u) \leq \mathbb{I}[u \leq \rho]$ for any $u \in \mathbb{R}$ (and in particular, for $u = \rho_h(x,y)$). □

[5] While these results assume a specific linear structure of the hypothesis class, it is easy to verify that the results hold in the more general setting described here.