Supplemental materials for Submission “Distributionally robust, skeptical inferences in multi-label problems”

Proof of Lemma 2 Let us first develop $\mathbb{E}\left[\ell_H(y^2,\cdot)-\ell_H(y^1,\cdot)|X=x\right]$:

$$\sum_{\boldsymbol{y}\in\mathscr{Y}}\left(\sum_{i=1}^{m}\mathbb{1}_{y_i\neq y_i^2}-\sum_{i=1}^{m}\mathbb{1}_{y_i\neq y_i^1}\right)P_x(Y=\boldsymbol{y}) \tag{20}$$

$$\sum_{y_1\in\{0,1\}}\sum_{y_2\in\{0,1\}}\cdots\sum_{y_m\in\{0,1\}}\left(\sum_{i=1}^{m}\mathbb{1}_{y_i\neq y_i^2}-\sum_{i=1}^{m}\mathbb{1}_{y_i\neq y_i^1}\right)P_x(Y=\boldsymbol{y}) \tag{21}$$

For a given $k\in\{1,\dots,m\}$, let us consider the rewriting

$$\overbrace{\sum_{y_1\in\{0,1\}}\sum_{y_2\in\{0,1\}}\cdots\sum_{y_m\in\{0,1\}}}^{m-1}\left[\sum_{y_k\in\{0,1\}}\left(\sum_{i=1}^{m}\mathbb{1}_{y_i\neq y_i^2}-\sum_{i=1}^{m}\mathbb{1}_{y_i\neq y_i^1}\right)\right]P_x(Y=\boldsymbol{y}). \tag{22}$$

Developing the sum between brackets, we get

$$\sum_{y_k\in\{0,1\}}\sum_{i=1}^{m}\mathbb{1}_{y_i\neq y_i^2}P_x(Y=\boldsymbol{y})-\sum_{y_k\in\{0,1\}}\sum_{i=1}^{m}\mathbb{1}_{y_i\neq y_i^1}P_x(Y=\boldsymbol{y})\quad\text{(by linearity)} \tag{23}$$

Developing again the left term, we obtain

$$\begin{aligned}\sum_{y_k\in\{0,1\}}\sum_{i=1}^{m}\mathbb{1}_{y_i\neq y_i^2}P_x(Y=\boldsymbol{y}) &= \sum_{y_k\in\{0,1\}}\left(\mathbb{1}_{y_1\neq y_1^2}+\mathbb{1}_{y_2\neq y_2^2}+\cdots+\mathbb{1}_{y_m\neq y_m^2}\right)P_x(Y=\boldsymbol{y})\\ &= \mathbb{1}_{y_1\neq y_1^2}\sum_{y_k\in\{0,1\}}P_x(Y^k=y_k)+\cdots+\sum_{y_k\in\{0,1\}}\mathbb{1}_{y_k\neq y_k^2}P_x(Y^k=y_k)+\\ &\quad\cdots+\mathbb{1}_{y_m\neq y_m^2}\sum_{y_k\in\{0,1\}}P_x(Y^k=y_k)\\ &= \mathbb{1}_{y_1\neq y_1^2}P_x(Y_{\{-k\}})+\cdots+\sum_{y_k\in\{0,1\}}\mathbb{1}_{y_k\neq y_k^2}P_x(Y^k=y_k)+\\ &\quad\cdots+\mathbb{1}_{y_m\neq y_m^2}P_x(Y_{\{-k\}})\\ &= \sum_{y_k\in\{0,1\}}\mathbb{1}_{y_k\neq y_k^2}P_x(Y^k=y_k)+\sum_{i=1,i\neq k}^{m}\mathbb{1}_{y_i\neq y_i^2}P_x(Y_{\{-k\}}),\end{aligned}$$

where $P_x(Y^k=y_k):=P_x(Y_1,\dots,Y_k=y_k,\dots,Y_m)$ and $P_x(Y_{\{-k\}}):=P_x(Y_1,\dots,Y_{k-1},Y_{k+1},\dots,Y_m)$.

Similarly, we get for the right term

$$\sum_{y_k\in\{0,1\}}\sum_{i=1}^{m}\mathbb{1}_{y_i\neq y_i^1}P_x(Y=\boldsymbol{y})=\sum_{y_k\in\{0,1\}}\mathbb{1}_{y_k\neq y_k^1}P_x(Y^k=y_k)+\sum_{i=1,i\neq k}^{m}\mathbb{1}_{y_i\neq y_i^1}P_x(Y_{\{-k\}})$$

We put back these rewritten sums in Equation (23)

$$\begin{aligned}&\overbrace{\sum_{y_1\in\{0,1\}}\sum_{y_2\in\{0,1\}}\cdots\sum_{y_m\in\{0,1\}}}^{m-1}\left[\sum_{y_k\in\{0,1\}}\mathbb{1}_{y_k\neq y_k^2}P_x(Y^k=y_k)-\sum_{y_k\in\{0,1\}}\mathbb{1}_{y_k\neq y_k^1}P_x(Y^k=y_k)\right.\\ &\quad\left.+\sum_{i=1,i\neq k}^{m}\mathbb{1}_{y_i\neq y_i^2}P_x(Y_{\{-k\}})-\sum_{i=1,i\neq k}^{m}\mathbb{1}_{y_i\neq y_i^1}P_x(Y_{\{-k\}})\right]=\\ &\sum_{\boldsymbol{y}\in\mathscr{Y}}(\mathbb{1}_{y_k\neq y_k^2}-\mathbb{1}_{y_k\neq y_k^1})P_x(Y=\boldsymbol{y})+\overbrace{\sum_{y_1\in\{0,1\}}\cdots\sum_{y_m\in\{0,1\}}}^{m-1}\left[\sum_{i=1,i\neq k}^{m}\mathbb{1}_{y_i\neq y_i^2}-\mathbb{1}_{y_i\neq y_i^1}\right]P_x(Y_{\{-k\}})\end{aligned} \tag{24}$$

The left term can be reduced in the following way:

$$\begin{aligned}\sum_{\mathbf{y}\in\mathscr{Y}}(\mathbb{1}_{y_k\neq y_k^2}-\mathbb{1}_{y_k\neq y_k^1})P_x(Y=\mathbf{y}) &= \sum_{\mathbf{y}\in\mathscr{Y}}\mathbb{1}_{y_k\neq y_k^2}P_x(Y=\mathbf{y})-\sum_{\mathbf{y}\in\mathscr{Y}}\mathbb{1}_{y_k\neq y_k^1}P_x(Y=\mathbf{y})\\ &= P_x(Y_k\neq y_k^2)-P_x(Y_k\neq y_k^1)\\ &= P_x(Y_k=y_k^1)-P_x(Y_k=y_k^2)\end{aligned}$$

since we have $P_x(Y_k\neq y_k)=1-P_x(Y_k=y_k)$. We can apply the same operations we just did on the right term of Equation (24), and do so recursively, to finally obtain

$$\sum_{i=1}^{m}P(Y_i=y_i^1)-P(Y_i=y_i^2) \tag{25}$$

■

Proof of Proposition 3 Using Equation (9), one can readily see that

$$\mathbf{y}^1\succ_M\mathbf{y}^2 \iff \inf_{P\in\mathscr{P}}\sum_{i\in\mathscr{I}_{\mathbf{y}^1\neq\mathbf{y}^2}}P(Y_i=y_i^1)-P(Y_i=y_i^2)>0 \tag{26}$$

$$\iff \inf_{P\in\mathscr{P}}\sum_{i\in\mathscr{I}}P(Y_i=a_i)-P(Y_i=\overline{a}_i)>0 \tag{27}$$

Accounting for the fact that $P(Y_i=a_i)+P(Y_i=\overline{a}_i)=1$, we get

$$\iff \inf_{P\in\mathscr{P}}\sum_{i\in\mathscr{I}}2P(Y_i=a_i)-1>0 \tag{28}$$

$$\iff \inf_{P\in\mathscr{P}}\sum_{i\in\mathscr{I}}P(Y_i=a_i)>\frac{|\mathscr{I}|}{2} \tag{29}$$

■

Proof of Proposition 4 First, let us simply notice that $P(Y_i=a_i)=\sum_{y\in\mathscr{Y}}\mathbb{1}_{y_i=a_i}p(Y=y)$ and $\mathbb{1}_{y_i=a_i}=\mathbb{1}_{y_i\neq\overline{a}_i}$. Putting these together, we get

$$\begin{aligned}\sum_{i\in\mathscr{I}}P(Y_i=a_i) &= \sum_{i\in\mathscr{I}}\sum_{y\in\mathscr{Y}}\mathbb{1}_{y_i\neq\overline{a}_i}P(Y=y)\\ &= \sum_{y\in\mathscr{Y}}\sum_{i\in\mathscr{I}}\mathbb{1}_{y_i\neq\overline{a}_i}P(Y=y) && \text{(by linearity)}\\ &= \mathbb{E}[\ell_H^*(\cdot,\overline{a}_{\mathscr{I}})]\end{aligned}$$

where $\ell_H^*(\cdot,\overline{a}_{\mathscr{I}})$ is the hamming loss calculated in the set of indices $\mathscr{I}=\{i_1,\dots,i_q\}$ of vector $\overline{a}_{\mathscr{I}}$, which is created in the line 5 of the Algorithm 1. Thus, we apply infimum, $\inf_{P\in\mathscr{P}}$, to each side of the last equation and get what we sought. ■

Proof of Proposition 5 Let us simply analyze the number of computations needed. We will need to perform m times the loop of Line 2. For a given i, we have that $\mathscr{Z}_i=\binom{m}{i}$, meaning that this is the number of elements to check in the loop starting Line 4. Finally, there 2^i elements to check in the loop starting Line 5. The table below summarise the different steps.

Index Line 2	$i=1$	$i=2$	$\dots$	$i=m-2$	$i=m-1$
$\lvert\mathscr{Z}_i\rvert$	$\frac{m!}{1!(m-1)!}$	$\frac{m!}{2!(m-2)!}$	$\cdots$	$\frac{m!}{(m-2)!2!}$	$\frac{m!}{(m-1)!1!}$
$\lvert\mathscr{Y}_z\rvert$	$\{0,1\}^1$	$\{0,1\}^2$	$\dots,$	$\{0,1\}^{m-2}$	$\{0,1\}^{m-1}$

Overall, the number of checks to perform amounts to

$$\sum_{k=1}^{m}2^{m-k}\frac{m!}{k!(m-k)!}=3^m-1 \tag{30}$$

■

The proofs of the next propositions, that concern partial binary vectors, require us to first prove an intermediate result characterising partial vectors in terms of the vector set they represent. More precisely, we first express a condition for a subset $\mathbb{Y}$ of $\mathscr{Y}$ to be a partial vector, in terms of its elements.

Lemma 10 *A subset $\mathbb{Y}$ belongs to the space $\mathscr{Y}^*$ if and only if*

$$\forall y, y' \in \mathbb{Y},\ \text{we have that all } y'' \in \mathscr{Y} \text{ s.t. } y''_i = y'_i \quad \forall i \in \mathscr{I}_{y=y'} \text{ are also in } \mathbb{Y}$$

Proof of Lemma 10 Only if: Immediate, since by assumption $\mathscr{I}_{y \neq y'} \subseteq \mathscr{I}^*$, the set of label indices on which we abstain.

If: Consider the set $D_{\mathbb{Y}} = \{j | \exists y, \mathbf{y'} \in \mathbb{Y}, y_j \neq y'_j\}$ of indices for which at least two elements of $\mathbb{Y}$ disagree. What we have to show is that under the condition of Lemma 10, any completion of $D_{\mathbb{Y}}$ is within $\mathbb{Y}$.

Without loss of generality, as we can always permute the indices, let us consider that $D_{\mathbb{Y}}$ are the $|D_{\mathbb{Y}}|$ first indices. We can then find a couple $y, y' \in \mathbb{Y}$ such that the k first elements are distinct, that is $\mathscr{I}_{y \neq y'} = \{1, \ldots, k\}$. It follows that the subset of vectors

$$(\underbrace{*, \ldots, *}_{k \text{ times}}, y_{k+1}, \ldots, y_{|D_{\mathbb{Y}}|}, y_{|D_{\mathbb{Y}}|+1}, \ldots, y_m) \tag{31}$$

is within $\mathbb{Y}$, by assumption. If $k < |D_{\mathbb{Y}}|$, we can find a vector y'' such that its k' next elements (after the kth first) are different from y, i.e., $y_j \neq y''_j$ for $j = k+1, \ldots, k+k'$ with $k+k' \leq |D_{\mathbb{Y}}|$. Note that $k' \geq 1$ by assumption. Since the vector (31) is in $\mathbb{Y}$, we can always consider the vector y such that its k first elements are different from those of y'', that is in $\mathbb{Y}$. Since $\mathscr{I}_{y \neq y'} = \{1, \ldots, k+k'\}$, the subset of vectors

$$(\underbrace{*, \ldots, *}_{k+k' \text{ times}}, y_{k+k'+1}, \ldots, y_{|D_{\mathbb{Y}}|}, y_{|D_{\mathbb{Y}}|+1}, \ldots, y_m)$$

is also in $\mathbb{Y}$. Since we can repeat this construction until having two vectors with the $|D_{\mathbb{Y}}|$ first labels different, this finishes the proof. ■

Proof of Proposition 8 We will first characterize the set of solutions that are Bayes optimal for some $p \in \mathscr{P}$, that is

$$\hat{\mathbb{Y}}^O_{\ell,\mathscr{P}} = \{\mathbf{y} \in \mathscr{Y} | \exists P \in \mathscr{P} \text{ s.t. } \mathbf{y} = \hat{\mathbf{y}}^P_\ell\}. \tag{32}$$

The reason for investigating this set is that it is known that it is an inner approximation of $\hat{\mathbb{Y}}^M_{\ell_H,\mathscr{P}_{BR}}$, that is $\hat{\mathbb{Y}}^O_{\ell,\mathscr{P}} \subseteq \hat{\mathbb{Y}}^M_{\ell_H,\mathscr{P}_{BR}}$. Given Equation (6), if we consider a set $\mathscr{P}_{BR}$ we have that

$$\mathbf{y} \in \hat{\mathbb{Y}}^O_{\ell_H,\mathscr{P}_{BR}} \iff \begin{cases} \underline{p}_i \leq 0.5 & \text{for } i \in \mathscr{I}_{y=0} \\ \overline{p}_i \geq 0.5 & \text{for } i \in \mathscr{I}_{y=1} \end{cases} \tag{33}$$

where $\mathscr{I}_{\mathbf{y}=0}$, $\mathscr{I}_{\mathbf{y}=1}$ are the indices of labels for which $y_i = 0$ and $y_i = 1$. Indeed, since here we start from the marginals, $\mathbf{y}$ is optimal according to Hamming loss and a distribution in $\mathscr{P}_{BR}$ iff we can fix p_i to be lower than 0.5 if $y_i = 0$, and higher else.

Now, let us consider two vectors $\mathbf{y}^1, \mathbf{y}^2$ and the indices $\mathscr{I}_{\mathbf{y}^1 \neq \mathbf{y}^2}$. Given the first part of this proof, if $\mathbf{y}^1, \mathbf{y}^2 \in \hat{\mathbb{Y}}^O_{\ell_H,\mathscr{P}}$, this means that $0.5 \in [\underline{p}_i, \overline{p}_i]$ for any $i \in \mathscr{I}_{\mathbf{y}^1 \neq \mathbf{y}^2}$. Therefore, given any vector $\mathbf{y}''$ such that $\mathbf{y}''_i = \mathbf{y}^1_i$ for $i \in \mathscr{I}_{\mathbf{y}^1 = \mathbf{y}^2}$, for the other indices $i \in \mathscr{I}_{\mathbf{y}^1 \neq \mathbf{y}^2}$, we can always fix a precise value $p_i \in [\underline{p}_i, \overline{p}_i]$ such that $\mathbf{y}''$ is also optimal w.r.t. p. More precisely, assume the assignments p^1_i and p^2_i result in $\mathbf{y}^1, \mathbf{y}^2$ being optimal predictions for the Hamming loss, respectively. Then $\mathbf{y}''$ is optimal for the assignment

$$p''_i = \begin{cases} p^1_i & \text{if } \mathbf{y}''_i = y^1_i \\ p^2_i & \text{if } y''_i = y^2_i \end{cases}$$

that is by definition within $\mathscr{P}_{BR}$. This means in particular that $\hat{\mathbb{Y}}^O_{\ell,\mathscr{P}} = \hat{\mathbf{y}}^*_{\ell_H,\mathscr{P}_{BR}}$.

To show that it also coincides with $\hat{\mathbf{y}}^*_{\ell_H,\mathscr{P}_{BR}} = \hat{\mathbb{Y}}^M_{\ell_H,\mathscr{P}_{BR}}$, we will consider the fact that $\hat{\mathbf{y}}^*_{\ell_H,\mathscr{P}_{BR}} \subseteq \hat{\mathbb{Y}}^M_{\ell_H,\mathscr{P}_{BR}}$, and will demonstrate that any vector outside $\hat{\mathbf{y}}^*_{\ell_H,\mathscr{P}_{BR}}$ is dominated (in the sense of Equation (3)) by a vector within $\hat{\mathbf{y}}^*_{\ell_H,\mathscr{P}_{BR}}$.

Let us consider a vector $\mathbf{y}' \notin \hat{\mathbf{y}}^*_{\ell_H,\mathscr{P}_{BR}}$, and the indices

$$\mathscr{I}_{\mathbf{y}' \neq \hat{\mathbf{y}}^*} = \{i : \hat{y}^*_{i,\ell_H,\mathscr{P}_{BR}} \neq *, \hat{y}^*_{i,\ell_H,\mathscr{P}_{BR}} \neq y'_i\}$$

on which they necessarily differ (as we can always set the labels for which $\hat{y}^*_{i,\ell_H,\mathscr{P}_{BR}} = *$ to be equal to y'_i). By Proposition 2, we have that

$$\hat{\mathbf{y}}^*_{\ell_H,\mathscr{P}_{BR}} \succ_M \mathbf{y}' \iff \inf_{P \in \mathscr{P}} \sum_{i \in \mathscr{I}_{\mathbf{y}' \neq \hat{\mathbf{y}}^*}} P(Y_i = \hat{y}^*_i) > \frac{|\mathscr{I}|}{2}$$

and since we have that $\hat{y}^*_i = 1 \Rightarrow \underline{P}(Y_i = 1) > 0.5$ and $\hat{y}^*_i = 0 \Rightarrow \underline{P}(Y_i = 0) > 0.5$, the right hand side inequality is satisfied. Hence, we can show that any vector outside $\hat{\mathbf{y}}^*_{\ell_H,\mathscr{P}_{BR}}$ is maximally dominated by another vector in $\hat{\mathbf{y}}^*_{\ell_H,\mathscr{P}_{BR}}$, meaning that $\hat{\mathbf{y}}^*_{\ell_H,\mathscr{P}_{BR}} \supseteq \hat{\mathbb{Y}}^M_{\ell_H,\mathscr{P}_{BR}}$. Combined with the fact that $\hat{\mathbf{y}}^*_{\ell_H,\mathscr{P}_{BR}} \subseteq \hat{\mathbb{Y}}^M_{\ell_H,\mathscr{P}_{BR}}$, this finishes the proof. ■