Supplementary Material for Unbiased Risk Estimators Can Mislead: A Case Study of Learning with Complementary Labels

A. Proofs

A.1. Proof of Proposition 1

Proof. Let $\boldsymbol{\eta}$ and $\overline{\boldsymbol{\eta}}$ denote the conditional distribution $\mathbb{P}(Y \mid X)$ and $\mathbb{P}(\overline{Y} \mid X)$ respectively, where $\boldsymbol{\eta}_k(x) = \mathbb{P}(Y = k \mid x)$ and $\overline{\boldsymbol{\eta}}_k(x) = \mathbb{P}(\overline{Y} = k \mid x)$. Since $\bar{y}$ only depends on y, we have $\overline{\boldsymbol{\eta}}(x) = T^\top \boldsymbol{\eta}(x)$. The unbiased risk estimator can be derived as follows:

$$
\begin{aligned}
R(\boldsymbol{g}; \ell) &= \mathbb{E}_{(x,y)\sim D}[\ell(y, \boldsymbol{g}(x))] = \mathbb{E}_X \mathbb{E}_{Y\sim\boldsymbol{\eta}(X)}[\ell(Y, \boldsymbol{g}(X))] \\
&= \mathbb{E}_X[\boldsymbol{\eta}(X)^\top \ell(\boldsymbol{g}(X))] = \mathbb{E}_X[\overline{\boldsymbol{\eta}}(X)^\top (T^{-1})\ell(\boldsymbol{g}(X))] \\
&= \mathbb{E}_{(x,\overline{y})\sim\overline{D}}[e_{\overline{y}}^\top (T^{-1})\ell(\boldsymbol{g}(x))]
\end{aligned}
$$

□

A.2. Proof of Proposition 2

Proof. Given the following two properties of ℓ_{01}:

$$
\begin{aligned}
&\sum_{i=1}^{K} \ell_{01}(i, \boldsymbol{g}(x)) = K - 1 \quad \text{and} \\
&\ell_{01}(\overline{y}, \boldsymbol{g}(x)) + \overline{\ell}_{01}(\overline{y}, \boldsymbol{g}(x)) = 1
\end{aligned}
$$

An unbiased risk estimator of classification error can be obtained by:

$$
\begin{aligned}
R(\boldsymbol{g}; \ell_{01}) &= \mathbb{E}_{(x,\overline{y})\sim\overline{D}}\left[-(K-1)\ell_{01}(\overline{y}, \boldsymbol{g}(x)) + \sum_{j=1}^{K} \ell_{01}(j, \boldsymbol{g}(x)) \right] \\
&= \mathbb{E}_{(x,\overline{y})\sim\overline{D}}\left[(K-1)(1 - \ell_{01}(\overline{y}, \boldsymbol{g}(x))) \right] \\
&= (K-1)\mathbb{E}_{(x,\overline{y})\sim\overline{D}}\left[\overline{\ell}_{01}(\overline{y}, \boldsymbol{g}(x)) \right] = (K-1)\overline{R}(\boldsymbol{g}; \overline{\ell}_{01})
\end{aligned}
$$

□

A.3. Proof of Proposition 3

Proof. The proposition can be derived by using the linearity of the gradient operator:

$$
\begin{aligned}
\mathbb{E}_{\overline{y}|y}\left[\nabla_\theta \overline{\ell}(\overline{y}, \boldsymbol{g}(x))\right] &= \nabla_\theta \mathbb{E}_{\overline{y}|y}\left[\overline{\ell}(\overline{y}, \boldsymbol{g}(x))\right] \\
&= \nabla_\theta \left[\frac{1}{K-1} \sum_{y' \neq y} \left[-(K-1)\ell(y', \boldsymbol{g}(x)) + \sum_{j=1}^{K} \ell(j, \boldsymbol{g}(x)) \right] \right] \\
&= \nabla_\theta \left[-\sum_{y' \neq y} \ell(y', \boldsymbol{g}(x)) + \sum_{j=1}^{K} \ell(j, \boldsymbol{g}(x)) \right] = \nabla_\theta \ell(y, \boldsymbol{g}(x))
\end{aligned}
$$

□