APPENDIX TO "RETHINKING CLASS-PRIOR ESTIMATION FOR POSITIVE-UNLABELED LEARNING"

A PROOFS

In this section, we show all the proofs.

A.1 PROOF OF PROPOSITION 1

Proposition 1. *Let* $\beta^* = \inf_{S\in\mathfrak{S},P_{\rm p}(S)>0}\frac{P_{\rm n}(S)}{P_{\rm p}(S)}$ *be the maximum proportion of* $P_{\rm p}$ *in* $P_{\rm n}$*, given* $P_{\rm u} = (1-\pi)P_{\rm n} + \pi P_{\rm p}$*, for* $0 < \pi \le 1$*, we have*

$$\kappa^* = \pi + (1-\pi)\inf_{S\in\mathfrak{S},P_{\rm p}(S)>0}\frac{P_{\rm n}(S)}{P_{\rm p}(S)} = \pi + (1-\pi)\beta^*. \tag{1}$$

Proof. Let κ^* be the maximum proportion of $P_{\rm p}$ in $P_{\rm u}$, which can be formulated as $\kappa^* = \inf_{S\in\mathfrak{S},P_{\rm p}(S)>0}\frac{P_{\rm u}(S)}{P_{\rm p}(S)}$. Then,

$$\begin{aligned}\kappa^* &= \inf_{S\in\mathfrak{S},P_{\rm p}(S)>0}\frac{(1-\pi)P_{\rm n}(S)+\pi P_{\rm p}(S)}{P_{\rm p}(S)}\\ &= \inf_{S\in\mathfrak{S},P_{\rm p}(S)>0}\frac{(1-\pi)P_{\rm n}(S)}{P_{\rm p}(S)}+\pi\\ &= \pi+(1-\pi)\inf_{S\in\mathfrak{S},P_{\rm p}(S)>0}\frac{P_{\rm n}(S)}{P_{\rm p}(S)}. \end{aligned}\tag{2}$$

By letting $\beta^* = \inf_{S\in\mathfrak{S},P_{\rm p}(S)>0}\frac{P_{\rm n}(S)}{P_{\rm p}(S)}$, $\kappa^* = \pi + (1-\pi)\beta^*$ which completes the proof. □

A.2 PROOF OF LEMMA 1

Lemma 1. *Let* M *be a probability measure over a measurable space* $(\mathcal{X},\mathfrak{S})$*. For any set* $A\in\mathfrak{S}$*, we have* $M^A + M^{A^c} = M$*.*

Proof. Let $A^c = \mathcal{X}\setminus A$. Let 2^A and 2^{A^c} be the power sets on A and A^c, respectively. According to Definition 2 in the main paper, M^A and M^{A^c} are defined as follows,

$$\forall S\in\mathfrak{S}, M^A(S) = M(S\cap A);$$
$$\forall S\in\mathfrak{S}, M^{A^c}(S) = M(S\cap A^c).$$

To prove $M = M^A + M^{A^c}$, we need to prove $\forall S\in\mathfrak{S}, M^A(S)+M^{A^c}(S) = M(S)$.

$\forall S\in\mathfrak{S}$,

$$\begin{aligned}M^A(S)+M^{A^c}(S) &= M(S\cap A)+M(S\cap A^c) = M((S\cap A)\cup(S\cap A^c))\\ &= M(S\cap(A\cup A^c)) = M(S\cap\mathcal{X}) = M(S),\end{aligned}$$

which completes the proof. □

A.3 Proof of Theorem 1

Theorem 1. *Let $P_{\mathrm{u}} = (1-\pi)P_{\mathrm{n}} + \pi P_{\mathrm{p}}$. Let $A \subset \mathrm{support}(P_{\mathrm{u}})$. By regrouping P_{n}^A to P_{p}, P_{u} can be written as a mixture, i.e., $P_{\mathrm{u}} = (1-\pi')P_{\mathrm{n}'} + \pi' P_{\mathrm{p}'}$, where*

$$\pi' = \pi + (1-\pi)P_{\mathrm{n}}(A), \tag{3}$$

$$P_{\mathrm{n}'} = \frac{P_{\mathrm{n}}^{A^c}}{P_{\mathrm{n}}(A^c)}, \;\; P_{\mathrm{p}'} = \frac{(1-\pi)P_{\mathrm{n}}^A + \pi P_{\mathrm{p}}}{(1-\pi)P_{\mathrm{n}}(A) + \pi}, \tag{4}$$

and $P_{\mathrm{n}'}$ and $P_{\mathrm{p}'}$ satisfy the anchor set assumption.

Proof. Firstly, we prove that by regrouping P_{n}^A to P_{p}, P_{u} is a convex combination of two new class-conditional distributions, i.e., $P_{\mathrm{u}} = (1-\pi')P_{\mathrm{n}'} + \pi' P_{\mathrm{p}'}$.

Let $A \in \mathfrak{S}$, we split P_{n} as $P_{\mathrm{n}}^{A^c}$ and P_{n}^A, transport P_{n}^A to P_{p} to regroup them together, i.e.,

$$P_{\mathrm{u}} = (1-\pi)P_{\mathrm{n}} + \pi P_{\mathrm{p}} = (1-\pi)(P_{\mathrm{n}}^A + P_{\mathrm{n}}^{A^c}) + \pi P_{\mathrm{p}} = (1-\pi)P_{\mathrm{n}}^{A^c} + ((1-\pi)P_{\mathrm{n}}^A + \pi P_{\mathrm{p}}). \tag{5}$$

Normalizing $P_{\mathrm{n}}^{A^c}$ and $((1-\pi)P_{\mathrm{n}}^A + \pi P_{\mathrm{p}})$ in Eq. (5) to probability measures, we have

$$\begin{aligned} P_{\mathrm{u}} &= (1-\pi)P_{\mathrm{n}}^{A^c} + ((1-\pi)P_{\mathrm{n}}^A + \pi P_{\mathrm{p}}) \\ &= ((1-\pi)P_{\mathrm{n}}^{A^c}(\mathcal{X}))\frac{P_{\mathrm{n}}^{A^c}}{P_{\mathrm{n}}^{A^c}(\mathcal{X})} + ((1-\pi)P_{\mathrm{n}}^A(\mathcal{X}) + \pi P_{\mathrm{p}}(\mathcal{X}))\frac{(1-\pi)P_{\mathrm{n}}^A + \pi P_{\mathrm{p}}}{(1-\pi)P_{\mathrm{n}}^A(\mathcal{X}) + \pi P_{\mathrm{p}}(\mathcal{X})} \\ &= ((1-\pi)P_{\mathrm{n}}^{A^c}(A^c))\frac{P_{\mathrm{n}}^{A^c}}{P_{\mathrm{n}}^{A^c}(A^c)} + (\pi + (1-\pi)P_{\mathrm{n}}^A(A))\frac{(1-\pi)P_{\mathrm{n}}^A + \pi P_{\mathrm{p}}}{(1-\pi)P_{\mathrm{n}}^A(A) + \pi} \\ &= ((1-\pi)P_{\mathrm{n}}(A^c))\frac{P_{\mathrm{n}}^{A^c}}{P_{\mathrm{n}}(A^c)} + (\pi + (1-\pi)P_{\mathrm{n}}(A))\frac{(1-\pi)P_{\mathrm{n}}^A + \pi P_{\mathrm{p}}}{(1-\pi)P_{\mathrm{n}}(A) + \pi}, \end{aligned} \tag{6}$$

where the last two qualities are obtained by the definition of P_{n}^A and $P_{\mathrm{n}}^{A^c}$. Let $P_{\mathrm{n}'} = \frac{P_{\mathrm{n}}^{A^c}}{P_{\mathrm{n}}(A^c)}$, $P_{\mathrm{p}'} = \frac{(1-\pi)P_{\mathrm{n}}^A + \pi P_{\mathrm{p}}}{(1-\pi)P_{\mathrm{n}}(A) + \pi}$ and $\pi' = \pi + (1-\pi)P_{\mathrm{n}}(A)$, then Eq. (6) becomes

$$P_{\mathrm{u}} = (1-\pi')P_{\mathrm{n}'} + \pi' P_{\mathrm{p}'},$$

which shows that P_{u} can be made to a convex combination of new class-conditional distributions $P_{\mathrm{n}'}$ and $P_{\mathrm{p}'}$ by regrouping P_{n}^A with P_{p}.

Now we prove that $P_{\mathrm{n}'}$ and $P_{\mathrm{p}'}$ satisfy the anchor set assumption by checking whether $P_{\mathrm{n}'}(A) = 0$ and $P_{\mathrm{p}'}(A) > 0$.

By the definition of P_{n} and $P_{\mathrm{n}}^{A^c}$, we have

$$P_{\mathrm{n}'}(A) = \frac{P_{\mathrm{n}}^{A^c}(A)}{P_{\mathrm{n}}(A^c)} = 0. \tag{7}$$

By the definition of P_{p} and P_{n}^A, we have

$$P_{\mathrm{p}'}(A) = \frac{(1-\pi)P_{\mathrm{n}}^A(A) + \pi P_{\mathrm{p}}(A)}{(1-\pi)P_{\mathrm{n}}(A) + \pi} = \frac{(1-\pi)P_{\mathrm{n}}(A) + \pi P_{\mathrm{p}}(A)}{(1-\pi)P_{\mathrm{n}}(A) + \pi} = \frac{P_{\mathrm{u}}(A)}{(1-\pi)P_{\mathrm{n}}(A) + \pi} > 0. \tag{8}$$

The last inequality holds because $A \subset \mathrm{support}(P_{\mathrm{u}})$. By combining Eq. (7) and Ineq. (8), we can conclude that $P_{\mathrm{n}'}$ and $P_{\mathrm{p}'}$ satisfy the anchor set assumption. □

A.4 Proof of Theorem 2

Theorem 2. *Let $P_{\mathrm{p}'}$ and $P_{\mathrm{n}'}$ be obtained by regrouping a set $A^* := \arg\min_{A \in \mathfrak{S}} \frac{P_{\mathrm{n}}(A)}{P_{\mathrm{p}}(A)}$*[1] *from P_{p} and P_{n}. 1). If P_{n} and P_{p} satisfy the irreduciblility assumption, then $\pi' = \pi$; 2). if P_{n} and P_{p} dissatisfy the irreduciblility assumption, then $\pi < \pi' < \pi + (1-\pi)\beta^* = \kappa^*$.*

[1] We have defined that the fraction tends to infinite if its numerator is larger than 0 and its denominator is 0. Additionally, the infimum may not be always exist, if it does not exist, we could use a sequence of sets that converges to the infimum value, but the convergence rate can be arbitrarily slow Scott (2015).

Proof. We define that a fraction tends to infinite if its numerator is larger than 0 and its denominator is 0. In this case, we could remove the constraint $P_{\mathrm{p}}(S) > 0$ in Eq. (2) and rewrite it to $\kappa^* = \pi + (1-\pi)\inf_{S \subseteq \mathrm{support}(P_{\mathrm{u}})} \frac{P_{\mathrm{n}}(S)}{P_{\mathrm{p}}(S)}$. We subtract it with the new class prior after regrouping (Eq. (3)), i.e.,

$$\begin{aligned}
\kappa^* - \pi' &= \pi + (1-\pi) \inf_{S \subseteq \mathrm{support}(P_{\mathrm{u}})} \frac{P_{\mathrm{n}}(S)}{P_{\mathrm{p}}(S)} - \pi - (1-\pi)P_{\mathrm{n}}(A^*) \\
&= (1-\pi)\left(\inf_{S \subseteq \mathrm{support}(P_{\mathrm{u}})} \frac{P_{\mathrm{n}}(S)}{P_{\mathrm{p}}(S)} - P_{\mathrm{n}}(A^*) \right). \qquad (9)
\end{aligned}$$

Not that $A^* := \arg\min_{A \in \mathfrak{S}} \frac{P_{\mathrm{n}}(A)}{P_{\mathrm{p}}(A)}$, if P_{n} is irreducible to P_{p}, $\inf_{S \subseteq \mathrm{support}(P_{\mathrm{u}})} \frac{P_{\mathrm{n}}(S)}{P_{\mathrm{p}}(S)} = 0$, so as $P_{\mathrm{n}}(A^*)$. Therefore $\kappa^* - \pi' = 0$ and $\pi' = \pi$.
If P_{n} is reducible to P_{p}, $P_{\mathrm{p}}(A^*) < 0$, then $\inf_{S \subseteq \mathrm{support}(P_{\mathrm{u}})} \frac{P_{\mathrm{n}}(S)}{P_{\mathrm{p}}(S)} > P_{\mathrm{n}}(A^*)$ and $\kappa^* - \pi' > 0$. Therefore $\pi < \pi'$ by Eq. (3), and $\pi < \pi' < \pi + (1-\pi)\beta = \kappa^*$. □

A.5 Proof of Theorem 3

For completeness, we illustrate the convergence property of ReCPE, which is presented by employing the estimator proposed by Blanchard et al. (2010).

Theorem 3. *Let $P_{\mathrm{u}} = (1-\pi)P_{\mathrm{n}} + \pi P_{\mathrm{p}}$. By selecting a set A and regrouping P_{n}^A to P_{p}. Then, with probability $1 - 2\delta$, the estimated class-prior $\hat{\pi}'$ based on solving $\inf_{S \in \mathfrak{S}, \hat{P}_{\mathrm{p}'}(S) > 0} \frac{\hat{P}_{\mathrm{x}}(S)}{\hat{P}_{\mathrm{p}'}(S)}$ satisfies*

$$|\hat{\pi}' - \pi| \leq \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})} + \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})} + (1-\pi)P_{\mathrm{n}}(A), \qquad (10)$$

where $\epsilon_{\delta,\mathcal{H}}(S) \triangleq 2\hat{\mathfrak{R}}_S(\mathcal{H}) + 3\sqrt{\frac{\log\frac{4}{\delta}}{2|S|}}$ for a set S, and $\hat{\mathfrak{R}}_S(\mathcal{H})$ is the empirical Rademacher complexity of $\mathcal{H}$.

Proof. Firstly, we illustrate Rademacher complexity bounds. Let $\mathcal{H}$ be a family of functions taking values in $\{-1,+1\}$, and let $\mathcal{D}$ be the distribution over the input space $\mathcal{X}$. Then, for any $\delta > 0$, with probability at least $1 - \delta/2$ over a sample $S = (x_1, \ldots, x_m)$ of size m drawn according to $\mathcal{D}$, for any function $h \in \mathcal{H}$,

$$R(h) - \hat{R}_S(h) \leq 2\hat{\mathfrak{R}}_S(\mathcal{H}) + 3\sqrt{\frac{\log\frac{4}{\delta}}{2m}}, \qquad (11)$$

where $R(h)$ is the expected risk of the function h, and $\hat{R}_S(h)$ is the empirical risk of the function h on the sample S (Mohri et al., 2018). Specifically, let c be a target concept, then,

$$R(h) = \mathop{\mathbb{E}}_{x \sim \mathcal{D}}[\mathbb{1}_{\{h(x_i) \neq c(x_i)\}}], \; \hat{R}_S(h) = \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}_{\{h(x_i) \neq c(x_i)\}}.$$

After regrouping P_{n}^A to P_{p} and creating $P_{\mathrm{p}'} = \frac{(1-\pi)P_{\mathrm{n}'}A + \pi P_{\mathrm{p}}}{(1-\pi)P_{\mathrm{n}'}(A) + \pi}$, P_{u} can be written as a mixture, i.e., $P_{\mathrm{u}} = (1-\pi')P_{\mathrm{n}'} + \pi' P_{\mathrm{p}'}$. Additionally, $P_{\mathrm{n}'}(A) = 0$ and $P_{\mathrm{p}'}(A) > 0$. Then,

$$P_{\mathrm{u}}(A) = (1-\pi')P_{\mathrm{n}'}(A) + \pi' P_{\mathrm{p}'}(A) = \pi' P_{\mathrm{p}'}(A). \qquad (12)$$

In order to bring in the Rademacher complexity bounds to the above equation, we have to connect both $P_{\mathrm{u}}(A)$ and $P_{\mathrm{p}'}(A)$ with the expected risk. Let's define a function $h \in \mathcal{H}$ which is an indicator of the anchor set A. That is, $\forall x \in \mathcal{X}$,

$$h(x) = \begin{cases} 1, & x \in A \\ -1, & x \notin A, \end{cases} \qquad (13)$$

By treating the sample i.i.d. drawn from the distribution P_{u} as positive, we can rewrite the $P_{\mathrm{u}}(A)$ as follows,

$$\begin{aligned} P_{\mathrm{u}}(A) &= \int_{x\in A} p_{\mathrm{u}}(x)dx = \int_{x\in\mathcal{X}} p_{\mathrm{u}}(x)\mathbb{1}_{\{h(x)=1\}}dx \\ &= 1 - \int_{x\in\mathcal{X}} p_{\mathrm{u}}(x)\mathbb{1}_{\{h(x)\neq 1\}}dx = 1 - \mathop{\mathbb{E}}_{x\sim P_{\mathrm{u}}}[\mathbb{1}_{\{h(x_i)\neq 1\}}] = 1 - R_1(h), \end{aligned}$$

where $R_1(h)$ represents the false negative risk of the function h.

Similarly, by treating the sample i.i.d. drawn from the distribution $P_{\mathrm{p}'}$ as negative, , we can rewrite the $P_{\mathrm{p}'}(A)$ as follows,

$$P_{\mathrm{p}'}(A) = \int_{x\in A} f_{P_{\mathrm{p}'}}(x)dx = \int_{x\in\mathcal{X}} f_{P_{\mathrm{p}'}}(x)\mathbb{1}_{\{h(x)\neq 0\}}dx = \mathop{\mathbb{E}}_{x\sim P_{\mathrm{p}'}}[\mathbb{1}_{\{h(x_i)\neq 0\}}] = R_0(h),$$

where $R_0(h)$ represents the false positive risk of the function h.

Suppose we have samples S_{u} and $S_{\mathrm{p}'}$ with sample sizes $|S_{\mathrm{u}}|$ and $|S_{\mathrm{p}'}|$ i.i.d. drawn from P_{u} and $P_{\mathrm{p}'}$, respectively. Let $\hat{P}_{\mathrm{x}}(A)$ and $\hat{P}_{\mathrm{p}'}(A)$ be the empirical version of $P_{\mathrm{u}}(A)$ and $P_{\mathrm{p}'}(A)$, which are defined uniformly over the training samples, that is,

$$\hat{P}_{\mathrm{x}}(A) = \frac{1}{|S_{\mathrm{u}}|}\sum_{x\in S_{\mathrm{u}}} \mathbb{1}_{\{h(x_i)=1\}} = 1 - \frac{1}{|S_{\mathrm{u}}|}\sum_{x\in S_{\mathrm{u}}} \mathbb{1}_{\{h(x_i)\neq 1\}} = 1 - \hat{R}_{1,S_{\mathrm{u}}}(h), \tag{14}$$

$$\hat{P}_{\mathrm{p}'}(A) = \frac{1}{|S_{\mathrm{p}'}|}\sum_{x\in S_{\mathrm{p}'}} \mathbb{1}_{h(x_i)\neq 0} = \hat{R}_{0,S_{\mathrm{p}'}}(h). \tag{15}$$

By Eq. (12), the estimated $\hat{\pi}'$ is

$$\hat{\pi}' = \frac{\hat{P}_{\mathrm{x}}(A)}{\hat{P}_{\mathrm{p}'}(A)}. \tag{16}$$

By using the Rademacher complexity bounds and union bound, with probability $1-\delta$, we have both

$$\begin{aligned} P_{\mathrm{u}}(A) &= 1 - R_1(h) \geq 1 - \hat{R}_{1,S_{\mathrm{u}}}(h) - \left(2\hat{\mathfrak{R}}_{S_{\mathrm{u}}}(\mathcal{H}) + 3\sqrt{\frac{\log\frac{4}{\delta}}{|S_{\mathrm{u}}|}}\right) \\ &\triangleq 1 - \hat{R}_{1,S_{\mathrm{u}}}(h) - \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}}), \end{aligned} \tag{17}$$

and

$$\begin{aligned} P_{\mathrm{p}'}(A) &= R_0(h) \leq \hat{R}_{0,S_{\mathrm{p}'}}(h) + 2\hat{\mathfrak{R}}_{S_{\mathrm{p}'}}(\mathcal{H}) + 3\sqrt{\frac{\log\frac{4}{\delta}}{|S_{\mathrm{p}'}|}} \\ &\triangleq \hat{R}_{0,S_{\mathrm{p}'}}(h) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'}). \end{aligned} \tag{18}$$

Substituting $P_{\mathrm{u}}(A)$ and $P_{\mathrm{p}'}(A)$ in Eq. (12) with Eq. (17) and Eq. (18), we have

$$1 - \hat{R}_{1,S_{\mathrm{u}}}(h) - \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}}) \leq \pi' P_{\mathrm{p}'}(A) \leq \pi'\left(\hat{R}_{0,S_{\mathrm{p}'}}(h) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})\right), \tag{19}$$

By Eq. (14) and Eq. (15), the above inequality can be rewritten as,

$$\hat{P}_{\mathrm{x}}(A) - \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}}) \leq \pi'\left(\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})\right).$$

Then we have that

$$
\begin{aligned}
\pi' &\geq \frac{\hat{P}_{\mathrm{x}}(A) - \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})} \\
&= \frac{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'}) - \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})} \frac{\hat{P}_{\mathrm{x}}(A) - \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}})}{\hat{P}_{\mathrm{p}'}(A)} \\
&= \left(1 - \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}\right)\left(\hat{\pi}' - \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}})}{\hat{P}_{\mathrm{p}'}(A)}\right) \\
&= \left(1 - \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}\right)\hat{\pi}' - \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})} \\
&= \hat{\pi}' - \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}\hat{\pi}' - \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})} \\
&\geq \hat{\pi}' - \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})} - \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}. \qquad (20)
\end{aligned}
$$

By the symmetric property of Eq. (10), with probability $1 - 2\delta$,

$$
|\hat{\pi}' - \pi'| \leq \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})} + \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}. \qquad (21)
$$

According Eq. (3), $\pi' = \pi + (1 - \pi)P_{\mathrm{n}}(A)$, then, with probability $1 - 2\delta$,

$$
|\hat{\pi}' - \pi| \leq \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})} + \frac{\epsilon_{\delta,\mathcal{H}}(S_{\mathrm{u}})}{\hat{P}_{\mathrm{p}'}(A) + \epsilon_{\delta,\mathcal{H}}(S_{\mathrm{p}'})} + (1 - \pi)P_{\mathrm{n}}(A).
$$

□

A.6 PROOF OF THEOREM 4

Theorem 4. *Let p_{u} and p_{p} be density functions of P_{u} and P_{p}, respectively. Let $q = P(C = 0)p_{\mathrm{u}} + P(C = 1)p_{\mathrm{p}}$. Let $\mathbb{1}_A : \mathcal{X} \to \{0, 1\}$ be the identity function which outputs 1 if $x \in \mathcal{X}$ is in the set A, and 0 otherwise. Then the set $A^* = \arg\min_{A \in \mathfrak{S}} \frac{\mathbb{E}_{x \sim q(X)}[\mathbb{1}_A(X=x)q(C=0|X=x)]}{\mathbb{E}_{x \sim q(X)}[\mathbb{1}_A(X=x)q(C=1|X=x)]}$.*

Recall that, in the main paper, we have defined another auxiliary distribution $q(X, C)$, where $C \in \{0, 1\}$ is the positive-vs-unlabeled label i.e., a class label distinguishing between the positive component and the whole mixture. Specifically, priors are $q(C = 1) := \frac{\pi}{1-\pi}$ and $q(C = 0) := \frac{1}{1-\pi}$; conditional densities are $q(X|C = 1) := P_{\mathrm{p}}$ and $q(X|C = 0) := P_{\mathrm{u}}$; class-posterior probabilities are $q(C = 0|X)$ and $q(C = 1|X)$.

Proof. Firstly, we prove that $\frac{P_{\mathrm{n}}(S)}{P_{\mathrm{p}}(S)}$ is proportional to $\frac{P_{\mathrm{u}}(S)}{P_{\mathrm{p}}(S)}$.

$$
\begin{aligned}
\frac{P_{\mathrm{u}}(S)}{P_{\mathrm{p}}(S)} &= \frac{(1 - \pi)P_{\mathrm{n}}(S) + \pi P_{\mathrm{p}}(S)}{P_{\mathrm{p}}(S)} \\
&= (1 - \pi)\frac{P_{\mathrm{n}}(S)}{P_{\mathrm{p}}(S)} + \pi.
\end{aligned}
$$

Since $\frac{1}{1-\pi}$ and $\frac{\pi}{1-\pi}$ are constants, then $\frac{P_{\mathrm{n}}(S)}{P_{\mathrm{p}}(S)}$ is proportional to $\frac{P_{\mathrm{u}}(S)}{P_{\mathrm{p}}(S)}$, which completes the first part of the proof.

Recall that, in the main paper, we have defined another auxiliary distribution $q(X, C)$, where $C \in \{0, 1\}$ is the positive-vs-unlabeled label i.e., a class label distinguishing between the positive component and the whole mixture. Specifically, priors are $P(C = 1) := q(C = 1) := \frac{\pi}{1-\pi}$ and $P(C = 0) := q(C = 0) := \frac{1}{1-\pi}$; conditional densities are $q(X|C = 1) := p_{\mathrm{p}}$ and $q(X|C = 0) := p_{\mathrm{u}}$; class-posterior probabilities are $q(C = 0|X)$ and $q(C = 1|X)$. We have

$$
\frac{P_{\mathrm{u}}(S)}{P_{\mathrm{p}}(S)} = \frac{\int_{x \in S} q(X = x|C = 0)dx}{\int_{x \in S} q(X = x|C = 1)dx} = \frac{\int_{x \in \mathcal{X}} \mathbb{1}_A(X = x)q(X = x|C = 0)dx}{\int_{x \in \mathcal{X}} \mathbb{1}_A(X = x)q(X = x|C = 1)dx}. \qquad (22)
$$

By using Bayesian rules, the above equation can be written as,

$$
\begin{aligned}
\frac{P_{\mathrm{u}}(S)}{P_{\mathrm{p}}(S)} &= \frac{\int_{x\in\mathcal{X}} \mathbb{1}_A(X=x)q(X=x|C=0)dx}{\int_{x\in\mathcal{X}} \mathbb{1}_A(X=x)q(X=x|C=1)dx} \\
&= \frac{P(C=1)}{P(C=0)} \frac{\int_{x\in\mathcal{X}} \mathbb{1}_A(X=x)q(C=0|X=x)q(x)dx}{\int_{x\in\mathcal{X}} \mathbb{1}_A(X=x)q(C=1|X=x)q(x)dx}. \\
&= \frac{P(C=1)}{P(C=0)} \frac{\mathbb{E}_{x\sim q(X)}[\mathbb{1}_A(X=x)q(C=0|X=x)]}{\mathbb{E}_{x\sim q(X)}[\mathbb{1}_A(X=x)q(C=1|X=x)]}.
\end{aligned}
$$

Since $\frac{P(C=1)}{P(C=0)}$ is a constant, then $\frac{P_{\mathrm{u}}(S)}{P_{\mathrm{p}}(S)}$ is proportional to $\frac{\mathbb{E}_{x\sim q(X)}[\mathbb{1}_A(X=x)q(C=0|X=x)]}{\mathbb{E}_{x\sim q(X)}[\mathbb{1}_A(X=x)q(C=1|X=x)]}$. Combining with the first part of the proof, i.e., $\frac{P_{\mathrm{n}}(S)}{P_{\mathrm{p}}(S)}$ is proportional to $\frac{P_{\mathrm{u}}(S)}{P_{\mathrm{p}}(S)}$, we can conclude that $\frac{P_{\mathrm{n}}(S)}{P_{\mathrm{p}}(S)}$ is proportional to $\frac{\mathbb{E}_{x\sim q(X)}[\mathbb{1}_A(X=x)q(C=0|X=x)]}{\mathbb{E}_{x\sim q(X)}[\mathbb{1}_A(X=x)q(C=1|X=x)]}$. By definition of $A^* := \arg\min_{A\in\mathfrak{S}} \frac{P_{\mathrm{n}}(A)}{P_{\mathrm{p}}(A)}$, then $A^* = \arg\min_{A\in\mathfrak{S}} \frac{\mathbb{E}_{x\sim q(X)}[\mathbb{1}_A(X=x)q(C=0|X=x)]}{\mathbb{E}_{x\sim q(X)}[\mathbb{1}_A(X=x)q(C=1|X=x)]}$, which completes the proof. □

A.7 PROOF OF PROPOSITION 2

Proposition 2. *Let* $P_{\tilde{\mathrm{p}}'} = \frac{P_{\mathrm{u}}^A + P_{\mathrm{p}}}{P_{\mathrm{u}}(A)+1}$ *and* $P_{\mathrm{u}}(A) < \epsilon$*. Then,* $\forall \epsilon > 0$ *and* $\forall S \in \mathfrak{S}$, $|P_{\mathrm{p}'}(S) - P_{\tilde{\mathrm{p}}'}(S)| \leq \mathcal{O}(\epsilon)$.

Proof. To prove $P_{\mathrm{p}'}$ is a good surrogate of $P_{\mathrm{p}'}$, we show that with the decreasing of $P_{\mathrm{u}}(A)$, the difference between $P_{\mathrm{p}'}$ and $P_{\tilde{\mathrm{p}}'}$ becomes smaller. Formally, let $P_{\mathrm{u}}(A) < \epsilon$. For all $\epsilon > 0$ and for all $S \in \mathfrak{S}$, $|P_{\mathrm{p}'}(S) - P_{\tilde{\mathrm{p}}'}(S)| \leq \mathcal{O}(\epsilon)$.

Note that the definitions of $P_{\mathrm{p}'}$ and $P_{\tilde{\mathrm{p}}'}$ are

$$
P_{\mathrm{p}'} = \frac{(1-\pi)P_{\mathrm{n}}^A + \pi P_{\mathrm{p}}}{(1-\pi)P_{\mathrm{n}}(A) + \pi}; P_{\tilde{\mathrm{p}}'} = \frac{P_{\mathrm{u}}^A + P_{\mathrm{p}}}{P_{\mathrm{u}}(A)+1}.
$$

We firstly start to prove that for all $\epsilon > 0$ and for all $S \in \mathfrak{S}$, $P_{\mathrm{p}'}(S) - P_{\tilde{\mathrm{p}}'}(S) \leq \mathcal{O}(\epsilon)$.

$$
\begin{aligned}
P_{\mathrm{p}'}(S) - P_{\tilde{\mathrm{p}}'}(S) &= \frac{(1-\pi)P_{\mathrm{n}}^A(S) + \pi P_{\mathrm{p}}(S)}{(1-\pi)P_{\mathrm{n}}(A) + \pi} - \frac{P_{\mathrm{u}}^A(S) + P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A)+1} \\
&= \frac{(P_{\mathrm{u}}^A(S) - \pi P_{\mathrm{p}}^A(S)) + \pi P_{\mathrm{p}}(S)}{(1-\pi)P_{\mathrm{n}}(A) + \pi} - \frac{P_{\mathrm{u}}^A(S) + P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A)+1} \\
&\leq \frac{P_{\mathrm{u}}^A(S) + \pi P_{\mathrm{p}}(S)}{\pi} - \frac{P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A)+1} \\
&\leq \frac{P_{\mathrm{u}}^A(A) + \pi P_{\mathrm{p}}(S)}{\pi} - \frac{P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A)+1} \\
&= \frac{P_{\mathrm{u}}(A) + \pi P_{\mathrm{p}}(S)}{\pi} - \frac{P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A)+1} \\
&= \frac{P_{\mathrm{u}}(A)^2 + P_{\mathrm{u}}(A) + \pi P_{\mathrm{p}}(S)P_{\mathrm{u}}(A) + \pi P_{\mathrm{p}}(S) - \pi P_{\mathrm{p}}(S)}{\pi P_{\mathrm{u}}(A) + \pi} \\
&= \frac{P_{\mathrm{u}}(A)(P_{\mathrm{u}}(A) + \pi P_{\mathrm{p}}(S))}{\pi P_{\mathrm{u}}(A) + \pi} \\
&\leq \frac{P_{\mathrm{u}}(A)(P_{\mathrm{u}}(A) + \pi P_{\mathrm{p}}(S))}{\pi} \\
&= \mathcal{O}(\epsilon). \qquad (23)
\end{aligned}
$$

We then prove that for all $\epsilon > 0$ and for all $S \in \mathfrak{S}$, $P_{\tilde{\mathrm{p}}'}(S) - P_{\mathrm{p}'}(S) \leq \mathcal{O}(\epsilon)$.

$$
\begin{aligned}
P_{\tilde{\mathrm{p}}'}(S) - P_{\mathrm{p}'}(S) &= \frac{P_{\mathrm{u}}^{A}(S) + P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A) + 1} - \frac{(1-\pi)P_{\mathrm{n}}^{A}(S) + \pi P_{\mathrm{p}}(S)}{(1-\pi)P_{\mathrm{n}}(A) + \pi} \\
&\leq \frac{P_{\mathrm{u}}^{A}(S) + P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A) + 1} - \frac{(1-\pi)P_{\mathrm{n}}^{A}(S) + \pi P_{\mathrm{p}}(S)}{(1-\pi)P_{\mathrm{n}}(A) + \pi P_{\mathrm{p}}(A) + \pi} \\
&\leq \frac{P_{\mathrm{u}}^{A}(S) + P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A) + 1} - \frac{\pi P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A) + \pi} \\
&\leq \frac{P_{\mathrm{u}}^{A}(A) + P_{\mathrm{p}}(S)}{1} - \frac{\pi P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A) + \pi} \\
&= \frac{P_{\mathrm{u}}(A) + P_{\mathrm{p}}(S)}{1} - \frac{\pi P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A) + \pi} \\
&= \frac{P_{\mathrm{u}}(A)^2 + \pi P_{\mathrm{u}}(A) + P_{\mathrm{p}}(S) P_{\mathrm{u}}(A) + \pi P_{\mathrm{p}}(S) - \pi P_{\mathrm{p}}(S)}{P_{\mathrm{u}}(A) + \pi} \\
&= \frac{P_{\mathrm{u}}(A)(P_{\mathrm{u}}(A) + \pi + P_{\mathrm{p}}(S))}{P_{\mathrm{u}}(A) + \pi} \\
&\leq \frac{P_{\mathrm{u}}(A)(P_{\mathrm{u}}(A) + \pi + P_{\mathrm{p}}(S))}{\pi} \\
&= \mathcal{O}(\epsilon). \qquad (24)
\end{aligned}
$$

By combining (23) and (24), for all $\epsilon > 0$ and for all $S \subseteq \mathcal{X}$, $|P_{\mathrm{p}'}(S) - P_{\tilde{\mathrm{p}}'}(S)| \leq \mathcal{O}(\epsilon)$, which completes the proof. □

B More Experimental Results

In this section, we provide more experimental results.

B.1 Estimation Errors on UCL Datasets

In Table 1, for each baseline method and its regrouped version, we report the average and variance of the absolute estimation errors and the p-values obtained by using Wilcoxon signed rank test. Note the, a small p-value reflects the error of the Regrouped-MPE is significantly smaller than the error of its baseline. The real-word datasets are downloaded from the UCL machine learning database[2].

	AM	ReAM	DPL	ReDPL	EN	ReEN	KM1	ReKM1	KM2	ReKM2	ROC	ReROC	RPG	ReRPG
adult (800)	**0.127**	0.13	0.122	**0.108**∗	0.316	**0.295**	0.255	**0.132**	0.164	**0.153**	0.176	**0.153**	0.135	**0.134**
	±0.005	±0.005	±0.006	±0.005	±0.005	±0.005	±0.051	±0.01	±0.009	±0.007	±0.01	±0.007	±0.004	±0.004
	$p = 0.413$		$p = \underline{0.036}$		$p = \underline{0.0}$		$p = \underline{0.0}$		$p = 0.182$		$p = 0.111$		$p = 0.16$	
adult (1600)	**0.122**	0.124	0.089*	0.089*	0.31	**0.29**	0.131	**0.091**	**0.12**	0.13	0.121	**0.095**	**0.123**	0.137
	±0.005	±0.004	±0.003	±0.003	±0.004	±0.005	±0.015	±0.008	±0.007	±0.006	±0.006	±0.005	±0.002	±0.004
	$p = 0.775$		$p = 0.485$		$p = \underline{0.0}$		$p = \underline{0.0}$		$p = 0.985$		$p = \underline{0.025}$		$p = 0.934$	
adult (3200)	0.105	**0.086**	**0.054**	0.057	0.297	**0.279**	0.054	**0.04**∗	**0.082**	0.089	0.089	**0.067**	**0.114**	0.128
	±0.003	±0.004	±0.001	±0.002	±0.003	±0.004	±0.001	±0.001	±0.003	±0.003	±0.005	±0.003	±0.002	±0.004
	$p = \underline{0.001}$		$p = 0.519$		$p = \underline{0.0}$		$p = \underline{0.0}$		$p = 0.879$		$p = \underline{0.009}$		$p = 0.98$	
avila (800)	0.168	**0.152**	**0.129**	0.147	0.447	**0.422**	0.105	**0.075**∗	0.104	**0.081**	0.263	**0.228**	0.119	**0.111**
	±0.011	±0.009	±0.005	±0.004	±0.004	±0.004	±0.007	±0.003	±0.004	±0.003	±0.011	±0.012	±0.007	±0.005
	$p = \underline{0.015}$		$p = 0.978$		$p = \underline{0.0}$		$p = \underline{0.0}$		$p = \underline{0.0}$		$p = \underline{0.024}$		$p = \underline{0.047}$	
avila (1600)	0.165	**0.132**	0.104	**0.084**	0.439	**0.418**	0.086	**0.076**∗	0.108	**0.092**	0.191	**0.16**	0.123	**0.121**
	±0.011	±0.01	±0.003	±0.003	±0.003	±0.003	±0.005	±0.004	±0.004	±0.003	±0.007	±0.01	±0.005	±0.005
	$p = \underline{0.0}$		$p = \underline{0.002}$		$p = \underline{0.0}$		$p = 0.133$		$p = \underline{0.005}$		$p = \underline{0.002}$		$p = 0.369$	
avila (3200)	0.156	**0.133**	**0.05**∗	0.061	0.436	**0.42**	0.092	**0.078**	0.112	**0.092**	0.131	**0.095**	**0.121**	0.122
	±0.012	±0.01	±0.001	±0.001	±0.002	±0.002	±0.005	±0.003	±0.007	±0.003	±0.005	±0.004	±0.005	±0.005
	$p = \underline{0.001}$		$p = 0.998$		$p = \underline{0.0}$		$p = 0.658$		$p = \underline{0.008}$		$p = \underline{0.0}$		$p = 0.601$	
bank (800)	**0.135**	0.158	**0.116**∗	0.132	0.282	**0.264**	0.356	**0.216**	0.266	**0.238**	0.163	**0.15**	**0.163**	0.185
	±0.011	±0.009	±0.004	±0.004	±0.013	±0.015	±0.086	±0.029	±0.036	±0.019	±0.004	±0.006	±0.01	±0.022
	$p = 0.992$		$p = 1.0$		$p = \underline{0.0}$		$p = \underline{0.0}$		$p = 0.088$		$p = 0.103$		$p = 0.995$	
bank (1600)	**0.117**	0.167	**0.087**∗	0.105	0.262	**0.244**	0.178	**0.128**	0.203	**0.198**	0.129	**0.118**	**0.157**	0.167
	±0.007	±0.015	±0.001	±0.002	±0.009	±0.01	±0.02	±0.013	±0.021	±0.015	±0.004	±0.005	±0.01	±0.011
	$p = 1.0$		$p = 1.0$		$p = \underline{0.0}$		$p = \underline{0.0}$		$p = 0.812$		$p = 0.119$		$p = 0.453$	

[2]UCL machine learning database.

bank (3200)	**0.104** 0.127 ±0.009 ±0.008 $p = 0.962$	**0.073∗** 0.091 ±0.002 ±0.002 $p = 1.0$	0.248 **0.237** ±0.007 ±0.008 $p = \underline{0.0}$	0.124 **0.09** ±0.008 ±0.004 $p = \underline{0.008}$	**0.15** 0.16 ±0.014 ±0.005 $p = 0.986$	**0.093** 0.106 ±0.003 ±0.003 $p = 0.947$	**0.159** 0.18 ±0.005 ±0.012 $p = 0.967$
card (800)	0.131 **0.127∗** ±0.007 ±0.007 $p = 0.71$	0.174 **0.161** ±0.007 ±0.009 $p = \underline{0.018}$	0.465 **0.444** ±0.029 ±0.03 $p = \underline{0.0}$	0.293 **0.176** ±0.041 ±0.013 $p = \underline{0.0}$	0.203 **0.158** ±0.025 ±0.015 $p = \underline{0.0}$	0.247 **0.233** ±0.019 ±0.021 $p = 0.207$	0.177 **0.155** ±0.013 ±0.011 $p = \underline{0.0}$
card (1600)	0.173 **0.14** ±0.009 ±0.009 $p = \underline{0.027}$	0.14 0.14 ±0.004 ±0.003 $p = 0.478$	0.459 **0.437** ±0.028 ±0.028 $p = \underline{0.0}$	0.19 **0.135** ±0.009 ±0.003 $p = \underline{0.0}$	0.159 **0.129** ±0.011 ±0.004 $p = \underline{0.003}$	0.194 **0.163** ±0.01 ±0.008 $p = 0.111$	0.126 **0.115∗** ±0.005 ±0.008 $p = \underline{0.0}$
card (3200)	0.164 **0.134** ±0.006 ±0.003 $p = \underline{0.009}$	0.127 **0.12** ±0.004 ±0.002 $p = 0.204$	0.455 **0.435** ±0.025 ±0.025 $p = \underline{0.0}$	0.161 **0.113** ±0.002 ±0.002 $p = \underline{0.0}$	0.142 **0.122** ±0.004 ±0.002 $p = \underline{0.0}$	0.159 **0.152** ±0.005 ±0.004 $p = 0.268$	0.11 **0.108∗** ±0.004 ±0.009 $p = 0.095$
covtype (800)	0.16 **0.123** ±0.01 ±0.006 $p = \underline{0.0}$	0.155 **0.151** ±0.006 ±0.005 $p = 0.255$	0.367 **0.343** ±0.003 ±0.004 $p = \underline{0.0}$	0.157 **0.142** ±0.011 ±0.009 $p = \underline{0.012}$	**0.122** 0.13 ±0.008 ±0.009 $p = 0.973$	0.291 **0.258** ±0.019 ±0.016 $p = \underline{0.027}$	0.116 **0.105∗** ±0.003 ±0.003 $p = \underline{0.003}$
covtype (1600)	0.12 **0.1∗** ±0.006 ±0.004 $p = \underline{0.004}$	0.132 **0.109** ±0.003 ±0.004 $p = \underline{0.002}$	0.364 **0.339** ±0.002 ±0.003 $p = \underline{0.0}$	0.116 **0.113** ±0.004 ±0.003 $p = 0.359$	**0.121** 0.123 ±0.005 ±0.005 $p = 0.768$	0.199 **0.161** ±0.014 ±0.01 $p = \underline{0.011}$	0.109 **0.108** ±0.003 ±0.003 $p = 0.257$
covtype (3200)	0.128 **0.09** ±0.003 ±0.003 $p = \underline{0.0}$	0.093 **0.083∗** ±0.003 ±0.002 $p = \underline{0.032}$	0.354 **0.334** ±0.001 ±0.002 $p = \underline{0.0}$	**0.097** 0.109 ±0.004 ±0.003 $p = 0.876$	**0.124** 0.128 ±0.003 ±0.004 $p = 0.825$	0.157 **0.113** ±0.009 ±0.004 $p = \underline{0.0}$	0.109 **0.107** ±0.003 ±0.003 $p = 0.154$
egg (800)	0.153 **0.106∗** ±0.011 ±0.007 $p = \underline{0.002}$	**0.218** 0.225 ±0.013 ±0.008 $p = 0.662$	0.505 0.505 ±0.005 ±0.006 $p = 0.433$	**0.173** 0.264 ±0.032 ±0.027 $p = 0.991$	**0.119** 0.131 ±0.007 ±0.008 $p = 0.789$	0.476 **0.396** ±0.022 ±0.03 $p = \underline{0.005}$	0.171 **0.124** ±0.02 ±0.009 $p = \underline{0.009}$
egg (1600)	0.137 **0.12** ±0.007 ±0.008 $p = 0.076$	**0.121** 0.142 ±0.006 ±0.005 $p = 0.992$	**0.486** 0.489 ±0.006 ±0.006 $p = 0.805$	0.234 **0.214** ±0.033 ±0.02 $p = \underline{0.018}$	0.116 **0.108∗** ±0.007 ±0.006 $p = \underline{0.047}$	0.315 **0.238** ±0.022 ±0.019 $p = \underline{0.002}$	0.151 **0.114** ±0.011 ±0.006 $p = \underline{0.0}$
egg (3200)	0.126 **0.113** ±0.006 ±0.006 $p = 0.117$	**0.057∗** 0.073 ±0.003 ±0.004 $p = 0.938$	**0.485** 0.489 ±0.012 ±0.011 $p = 0.958$	0.26 **0.193** ±0.02 ±0.017 $p = \underline{0.0}$	0.134 **0.113** ±0.007 ±0.006 $p = \underline{0.0}$	0.163 **0.139** ±0.009 ±0.008 $p = \underline{0.015}$	0.142 **0.102** ±0.008 ±0.005 $p = \underline{0.0}$
magic04 (800)	0.099 **0.077** ±0.006 ±0.004 $p = \underline{0.012}$	0.072 **0.071** ±0.003 ±0.002 $p = 0.357$	0.312 **0.296** ±0.003 ±0.004 $p = \underline{0.0}$	0.111 **0.1** ±0.005 ±0.006 $p = \underline{0.0}$	0.071 **0.064** ±0.002 ±0.001 $p = 0.056$	0.141 **0.124** ±0.01 ±0.007 $p = 0.181$	0.055 **0.054∗** ±0.001 ±0.001 $p = 0.203$
magic04 (1600)	0.071 **0.056** ±0.002 ±0.002 $p = \underline{0.001}$	0.044 **0.043∗** ±0.002 ±0.001 $p = 0.497$	0.292 **0.274** ±0.002 ±0.002 $p = \underline{0.0}$	0.084 **0.072** ±0.003 ±0.004 $p = \underline{0.0}$	0.079 **0.065** ±0.003 ±0.002 $p = \underline{0.0}$	0.1 **0.073** ±0.004 ±0.003 $p = \underline{0.002}$	0.058 **0.052** ±0.001 ±0.001 $p = \underline{0.003}$
magic04 (3200)	0.069 **0.054** ±0.002 ±0.001 $p = \underline{0.0}$	**0.035∗** 0.036 ±0.001 ±0.002 $p = 0.562$	0.274 **0.258** ±0.001 ±0.001 $p = \underline{0.0}$	0.07 **0.047** ±0.003 ±0.002 $p = \underline{0.0}$	0.085 **0.063** ±0.002 ±0.002 $p = \underline{0.0}$	0.065 **0.047** ±0.003 ±0.002 $p = \underline{0.007}$	0.054 **0.052** ±0.001 ±0.001 $p = 0.176$
robot (800)	**0.053** 0.062 ±0.004 ±0.002 $p = 0.961$	0.049 **0.047∗** ±0.002 ±0.001 $p = 0.681$	0.19 **0.187** ±0.001 ±0.001 $p = 0.101$	0.232 **0.215** ±0.023 ±0.02 $p = 0.108$	**0.111** 0.114 ±0.007 ±0.007 $p = 0.975$	**0.119** 0.144 ±0.006 ±0.004 $p = 0.986$	**0.077** 0.084 ±0.003 ±0.003 $p = 0.838$
robot (1600)	0.053 **0.038∗** ±0.005 ±0.001 $p = 0.129$	0.087 **0.054** ±0.007 ±0.002 $p = \underline{0.0}$	0.139 **0.132** ±0.001 ±0.001 $p = \underline{0.0}$	0.15 **0.141** ±0.018 ±0.015 $p = \underline{0.003}$	**0.098** 0.099 ±0.005 ±0.005 $p = 0.849$	0.08 **0.075** ±0.004 ±0.002 $p = 0.477$	**0.076** 0.079 ±0.002 ±0.003 $p = 0.762$
robot (3200)	0.052 **0.039∗** ±0.003 ±0.002 $p = \underline{0.001}$	0.156 **0.119** ±0.01 ±0.007 $p = \underline{0.0}$	0.091 **0.085** ±0.0 ±0.0 $p = \underline{0.0}$	0.079 **0.077** ±0.007 ±0.006 $p = 0.161$	0.084 0.084 ±0.004 ±0.004 $p = 0.395$	0.063 **0.043** ±0.004 ±0.001 $p = 0.057$	**0.06** 0.066 ±0.002 ±0.003 $p = 0.988$
shuttle (800)	0.083 **0.031** ±0.039 ±0.001 $p = 0.271$	**0.016∗** 0.02 ±0.0 ±0.001 $p = 0.898$	0.041 **0.035** ±0.001 ±0.0 $p = \underline{0.0}$	**0.058** 0.083 ±0.002 ±0.003 $p = 1.0$	**0.035** 0.065 ±0.001 ±0.005 $p = 1.0$	**0.042** 0.047 ±0.001 ±0.002 $p = 0.699$	**0.035** 0.051 ±0.001 ±0.003 $p = 1.0$
shuttle (1600)	0.09 **0.045** ±0.048 ±0.011 $p = 0.958$	**0.011∗** 0.018 ±0.0 ±0.001 $p = 0.927$	0.04 **0.034** ±0.0 ±0.0 $p = \underline{0.0}$	**0.048** 0.079 ±0.001 ±0.003 $p = 1.0$	**0.024** 0.05 ±0.0 ±0.003 $p = 1.0$	**0.029** 0.043 ±0.001 ±0.003 $p = 0.913$	**0.026** 0.039 ±0.001 ±0.002 $p = 1.0$
shuttle (3200)	0.076 **0.028** ±0.039 ±0.0 $p = 0.949$	**0.012∗** 0.021 ±0.0 ±0.001 $p = 1.0$	0.043 **0.038** ±0.0 ±0.001 $p = \underline{0.004}$	**0.046** 0.07 ±0.001 ±0.002 $p = 1.0$	**0.018** 0.03 ±0.0 ±0.001 $p = 0.999$	**0.038** 0.045 ±0.005 ±0.004 $p = 0.811$	**0.028** 0.042 ±0.001 ±0.002 $p = 1.0$
average	0.116 **0.1** ±0.012 ±0.007 $p = \underline{0.0}$	0.094 **0.092∗** ±0.006 ±0.006 $p = 0.279$	0.311 **0.297** ±0.026 ±0.026 $p = \underline{0.0}$	0.146 **0.121** ±0.022 ±0.012 $p = \underline{0.0}$	0.117 **0.111** ±0.01 ±0.008 $p = \underline{0.0}$	0.157 **0.136** ±0.018 ±0.014 $p = \underline{0.0}$	0.106 **0.105** ±0.007 ±0.008 $p = \underline{0.002}$

Table 1: The first column provides the names of the datasets and the sample lengths. We bold the smaller average estimation errors by comparing each baseline method with its regrouped version. The smallest average estimation error among all methods in each row is highlighted with $*$. p-values are obtained by using the one-sided Wilcoxon signed rank test. We underline the p-values which are smaller than the 0.05 significant level. The last column is calculated by averaging trials on all the different datasets. The proposed Regrouping method provides significantly more accurate estimations than all the baseline.

REFERENCES

Gilles Blanchard, Gyemin Lee, and Clayton Scott. Semi-supervised novelty detection. *Journal of Machine Learning Research*, 11(Nov):2973–3009, 2010.

Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.

Clayton Scott. A rate of convergence for mixture proportion estimation, with application to learning from noisy labels. In *Artificial Intelligence and Statistics*, pp. 838–846, 2015.