

1 PROOFS OMITTED FROM THE MAIN TEXT

Proof of Theorem 11. First we define two auxiliary estimators $\tilde{A}_\Upsilon^{(q)}$ and $\tilde{A}_\Gamma^{(q)}$. Let $\mathbf{Y}^{(q)}$ be a vector of $\lfloor n/k \rfloor \cdot q$ size- k subsets of \mathcal{C}_Υ where the subsets of \mathcal{C}_Υ in each of the q non-overlapping size- $\lfloor n/k \rfloor$ segments $\mathbf{Y}_1^{(q)}, \mathbf{Y}_2^{(q)}, \dots, \mathbf{Y}_q^{(q)}$ of $\mathbf{Y}^{(q)}$ are sampled in the same way as the elements of the vector \mathbf{Y} in Lemma 3, all with the same \mathcal{C}_Υ (i.e. $\mathbf{Y}^{(q)}$ is the concatenation of the vectors $\mathbf{Y}_1^{(q)}, \mathbf{Y}_2^{(q)}, \dots, \mathbf{Y}_q^{(q)}$). Another vector $\mathbf{Z}^{(q)}$ which contains $\lfloor u/k \rfloor \cdot q$ size- k subsets of \mathcal{C}_Γ is sampled in the same way. Note that $\mathbf{Z}^{(q)}$ is independent of $\mathbf{Y}^{(q)}$. Let us define

$$\tilde{A}_\Upsilon^{(q)} = \frac{1}{q \cdot \lfloor n/k \rfloor} \sum_{S \in \mathbf{Y}^{(q)}} \mathbb{1}(\Upsilon \langle S \rangle \models \alpha) \text{ and,}$$

$$\tilde{A}_\Gamma^{(q)} = \frac{1}{q \cdot \lfloor u/k \rfloor} \sum_{S \in \mathbf{Z}^{(q)}} \mathbb{1}(\Gamma \langle S \rangle \models \alpha).$$

We can rewrite them as

$$\tilde{A}_\Upsilon^{(q)} = \frac{1}{q} \sum_{i=1}^q \frac{1}{\lfloor n/k \rfloor} \sum_{S \in \mathbf{Y}_i^{(q)}} \mathbb{1}(\Upsilon \langle S \rangle \models \alpha),$$

$$\tilde{A}_\Gamma^{(q)} = \frac{1}{q} \sum_{i=1}^q \frac{1}{\lfloor u/k \rfloor} \sum_{S \in \mathbf{Z}_i^{(q)}} \mathbb{1}(\Gamma \langle S \rangle \models \alpha).$$

Let us denote $m_1 = \lfloor n/k \rfloor$, $m_2 = \lfloor u/k \rfloor$ and $T_i := \frac{1}{\lfloor n/k \rfloor} \sum_{S \in \mathbf{Y}_i^{(q)}} (\mathbb{1}(\Upsilon \langle S \rangle \models \alpha) - A_\aleph) - \frac{1}{\lfloor u/k \rfloor} \sum_{S \in \mathbf{Z}_i^{(q)}} (\mathbb{1}(\Gamma \langle S \rangle \models \alpha) - A_\aleph)$ (we note that $\mathbb{E}[T_i] = 0$). Using the same arguments as in the proof of Theorem 10, we obtain the following:

$$\begin{aligned} P[\tilde{A}_\Upsilon^{(q)} - \tilde{A}_\Gamma^{(q)} \geq \varepsilon] &\leq \sum_{i=1}^q \frac{1}{q} \cdot \mathbb{E}[\exp(h(T_i - \varepsilon))] \\ &\leq e^{-h\varepsilon} \exp\left(\frac{h^2}{8m_1}\right) \exp\left(\frac{h^2}{8m_2}\right) \\ &= \exp\left(-h\varepsilon + \frac{m_1 + m_2}{8m_1m_2} \cdot h^2\right) \end{aligned}$$

The bound achieves its minimum at $h = \frac{4\varepsilon m_1 m_2}{m_1 + m_2}$.

Thus, we get

$$P[\tilde{A}_\Upsilon^{(q)} - \tilde{A}_\Gamma^{(q)} \geq \varepsilon] \leq \exp\left(\frac{-2\varepsilon^2}{1/\lfloor n/k \rfloor + 1/\lfloor u/k \rfloor}\right),$$

symmetrically also $P[\tilde{A}_\Gamma^{(q)} - \tilde{A}_\Upsilon^{(q)} \geq \varepsilon] \leq \exp\left(\frac{-2\varepsilon^2}{1/\lfloor n/k \rfloor + 1/\lfloor u/k \rfloor}\right)$, and, using union bound, we get

$$P[|\tilde{A}_\Upsilon^{(q)} - \tilde{A}_\Gamma^{(q)}| \geq \varepsilon] \leq 2 \exp\left(\frac{-2\varepsilon^2}{1/\lfloor n/k \rfloor + 1/\lfloor u/k \rfloor}\right).$$

It follows from the strong law of large numbers (which holds for any Υ and Γ) that $P[\lim_{q \rightarrow \infty} \tilde{A}_\Upsilon^{(q)} = \hat{A}_\Upsilon \text{ and } \tilde{A}_\Gamma^{(q)} = \hat{A}_\Gamma] = 1$. Since q was arbitrary, the statement of the proposition follows. \square

2 REPRESENTING CONSTANTS USING AUXILIARY PREDICATES

In this paper we restricted ourselves to reasoning with theories that do not contain any constants. It is straightforward to extend our results to provide PAC-type bounds also for theories with constants by introducing auxiliary predicates. For instance, in the smokers domain, if we want to express that friends of Alice do not smoke, i.e. $\forall X : fr(alice, X) \Rightarrow \neg sm(X)$, then we may introduce an auxiliary predicate *friendOfAlice*/1 and the rule becomes $\forall X : friendOfAlice(X) \Rightarrow \neg sm(X)$. We note here that it is not necessary to add auxiliary predicates explicitly in practice. We use auxiliary predicates just for theoretical purposes to explain how the results about PAC-reasoning derived in this paper can be applied when constants are allowed.

This also reveals interesting properties of the problem. For instance, in order to do non-trivial reasoning based on k -entailment with a theory consisting only of the rule

$$\forall X, Y : sm(X) \wedge fr(X, Y) \Rightarrow sm(Y)$$

we need $k \geq 2$. However, for the rule

$$\forall X : friendOfAlice(X) \Rightarrow \neg sm(X)$$

we only need $k \geq 1$. Hence, for the derived PAC bounds, we can see that the expected number of errors made when using only the second rule grows as in the attribute-value case whereas the expected number of errors for the first rule may grow more quickly with the increasing size of the test examples (cf. Theorem 14).