Exercise 1

$\boxed{\Longrightarrow}$ First, let's show that the first statement implies the second. To do that, assume the first statement is true.

If we look into the definition of $\lim_{m\to\infty}$, then we will see that what we actually want to prove is that for every ε, $\exists\, m(\varepsilon)$, s.t. for all $m\in\mathbb{N}$ greater than $m(\varepsilon)$, it is true that $\mathbb{E}_{S\sim D^m}[L_D(A(S))]<\varepsilon$.

~~Since we assume that the first statement is true, consider m(~~

Let's fix ~~every~~ some $\varepsilon>0$. Since we assume that the first statement is true, consider $m(\frac{\varepsilon}{2},\frac{\varepsilon}{2})$, s.t. $\forall\, m\geq m(\frac{\varepsilon}{2},\frac{\varepsilon}{2})$, $\mathbb{P}_{S\sim D^m}[L_D(A(S))>\frac{\varepsilon}{2}]<\frac{\varepsilon}{2}$.

Let $\mathbb{E}_{S\sim D^m}[L_D(A(S))\,|\,L_D(A(S))<\varepsilon]$ denote the expectation of all $L_D(A(S))$ smaller than ε over samples S of size m. Then, $\forall\, m\geq m(\frac{\varepsilon}{2},\frac{\varepsilon}{2})$:

$$\mathbb{E}_{S\sim D^m}[L_D(A(S))]=\mathbb{P}_{S\sim D^m}\left(L_D(A(S))\leq\tfrac{\varepsilon}{2}\right)\underbrace{\mathbb{E}_{S\sim D^m}\left[L_D(A(S))\,|\,L_D(A(S))\leq\tfrac{\varepsilon}{2}\right]}_{\leq\frac{\varepsilon}{2}}+\mathbb{P}_{S\sim D^m}\left(L_D(A(S))\geq\tfrac{\varepsilon}{2}\right)\cdot$$

$$\cdot\,\mathbb{E}\left[L_D(A(S))\,|\,L_D(A(S))\geq\tfrac{\varepsilon}{2}\right]<$$

$$<\mathbb{P}\left(L_D(A(S))\leq\tfrac{\varepsilon}{2}\right)\cdot\tfrac{\varepsilon}{2}+\tfrac{\varepsilon}{2}\cdot\mathbb{E}\left[L_D(A(S))\,|\,L_D(A(S))>\tfrac{\varepsilon}{2}\right]\leq$$

$$\leq\tfrac{\varepsilon}{2}+\tfrac{\varepsilon}{2}=$$ (since both probability and the loss function are not greater than 1)

$$=\varepsilon.$$

We have shown that $\forall\,\varepsilon>0$, $\exists\, m(\varepsilon)$ s.t. $\forall\, m>m(\varepsilon)$, $m\in\mathbb{N}$, it is true that $\mathbb{E}_{S\sim D^m}[L_D(A(S))]<\varepsilon$

Hence, $\lim_{m\to\infty}\mathbb{E}_{S\sim D^m}[L_D(A(S))]=0$.

$\boxed{\Leftarrow}$ Now, let's assume that the second statement is true and show that it implies that the first statement is also true.

Let's fix some $\varepsilon, \delta > 0$. Now, consider $m(\varepsilon, \delta)$, s.t. $\forall\, m \geq m(\varepsilon,\delta), m \in \mathbb{N}$, $\mathbb{E}_{S \sim \mathcal{D}^m}[L_{\mathcal{D}}(A(S))] < \varepsilon\delta$ (we can do this because we assume that $\lim_{m \to \infty} \mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S))] = 0$)

Then, for all $m \geq m(\varepsilon, \delta)$:

$$\varepsilon\delta > \mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S))] =$$

$$= \mathbb{P}_{S\sim\mathcal{D}^m}(L_{\mathcal{D}}(A(S)) \leq \varepsilon)\, \mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S)) \mid L_{\mathcal{D}}(A(S)) \leq \varepsilon] + \mathbb{P}_{S\sim\mathcal{D}^m}(L_{\mathcal{D}}(A(S)) > \varepsilon)\, \mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S)) \mid L_{\mathcal{D}}(A(S)) > \varepsilon] >$$

$$> \mathbb{P}_{S\sim\mathcal{D}^m}(L_{\mathcal{D}}(A(S)) > \varepsilon)\, \mathbb{E}[L_{\mathcal{D}}(A(S)) \mid L_{\mathcal{D}}(A(S)) > \varepsilon]$$ (since probability ≥ 0) ~~is greater or equal smaller~~

$$> \mathbb{P}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S)) > \varepsilon] \cdot \varepsilon$$

Hence, $\forall\, m \geq m(\varepsilon, \delta)$, $\delta > \mathbb{P}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S)) > \varepsilon]$

We have shown that if the 2nd statement is correct, than $\forall\, \varepsilon, \delta > 0$, $\exists\, m(\varepsilon, \delta)$, s.t. $\forall\, m \geq m(\varepsilon,\delta)$, $\mathbb{P}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S)) > \varepsilon] < \delta$.

~~Since~~ From $\boxed{\Rightarrow}$ and $\boxed{\Leftarrow}$, we get that both statements are equivalent.

q. e. d.

Exercise 2

Fix some $\varepsilon, \delta > 0$. We want to show that $\exists\, m_{\mathcal{H}}(\varepsilon,\delta)$, s.t. for all $m \geq m_{\mathcal{H}}(\varepsilon,\delta)$ it holds that

$$\mathcal{D}^m(\{S : \forall h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| \leq \varepsilon\}) \geq 1-\delta$$

and that such $m_{\mathcal{H}}(\varepsilon,\delta)$ is ~~is smaller than~~ is smaller or equal to $\left\lceil \frac{2\log\left(\frac{2|\mathcal{H}|}{\delta}\right)(b-a)^2}{\varepsilon^2} \right\rceil$

$\mathcal{D}^m(\{S : \forall h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| \leq \varepsilon\}) \geq 1-\delta$ is equivalent to $\mathcal{D}^m(\{S : \exists h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| > \varepsilon\}) \leq \delta$.

Note that:

$$\mathcal{D}^m(\{S : \exists h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| > \varepsilon\}) =$$

$$= \mathcal{D}^m\Big(\bigcup_{h \in \mathcal{H}} \{S : |L_S(h) - L_{\mathcal{D}}(h)| > \varepsilon\}\Big) \leq$$

$$\leq \sum_{h \in \mathcal{H}} \mathcal{D}^m(\{S : |L_S(h) - L_{\mathcal{D}}(h)| > \varepsilon\}) \quad \text{(Union Bound Lemma)}$$

$$= \sum_{h \in \mathcal{H}} \mathbb{P}_{S \sim \mathcal{D}^m}\Big[\Big|\frac{1}{m}\sum_{i=1}^m \ell(h, z_i) - \mu\Big| > \varepsilon\Big] \quad (\text{where } z_i \in S = (z_1, \dots z_m))$$

$$\leq \sum_{h \in \mathcal{H}} 2\exp(-2m\varepsilon^2/(b-a)^2) \quad \text{(using Hoeffding's Inequality)}$$

$$= 2|\mathcal{H}| \exp\left(\frac{-2m\varepsilon^2}{(b-a)^2}\right) \leftarrow \text{we want this to be } \leq \delta$$

Therefore, if we choose $m \geq \left\lceil \frac{\log(2|H|/\delta)(b-a)^2}{\varepsilon^2} \right\rceil$,

$\mathcal{D}^m(\{S : \forall h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| \leq \varepsilon\})$ is going to be greater or equal to $1-\delta$

q.e.d.

$\boxed{\text{Exercise 3}}$

We want to prove that:

$$\mathbb{E}_{S\sim D^m}[L_D(ERM_{\mathcal{H}}(S)) - L_S(ERM_{\mathcal{H}}(S))] \geq 0$$

By linearity of expectation this is equivalent to:

$$\mathbb{E}_{S\sim D^m}[L_D(ERM_{\mathcal{H}}(S))] \geq \mathbb{E}_{S\sim D^m}[L_S(ERM_{\mathcal{H}}(S))]$$

Note that $\forall h \in \mathcal{H}$:

$$\mathbb{E}_{S\sim D^m}[L_S(h)] = \mathbb{E}_{S\sim D^m}[\frac{1}{m}\sum_{i=1}^{m} \ell(h, z_i)] \quad (\text{where } z_i \in S = \{z_1, \dots z_m\})$$

$$= \mathbb{E}_{S\sim D^m}[\ell(h, z_i)] \quad (\text{by linearlity of expectation})$$

$$= \mathbb{E}_{z\sim D}[\ell(h, z)] = L_D(h)$$

Consider a random sample S_j. Then for that sample it is true that:

$$L_D[ERM_{\mathcal{H}}(S_j)] = \mathbb{E}_{S\sim D^m}[L_S(ERM(S_j))] \geq \mathbb{E}_{S\sim D^m}[L_S(ERM(S))]$$

(since for a random sample S_k, $L_{S_k}(ERM(S_j)) \geq L_{S_k}(ERM(S_k))$)

~~Theref~~

Since this holds for all samples from D^m, we get that: it also holds for expected value of those samples:

$$\mathbb{E}_{S\sim D^m}[L_D(ERM_{\mathcal{H}}(S))] \geq \mathbb{E}_{S\sim D^m}[L_S(ERM_{\mathcal{H}}(S))]$$

q.e.d.