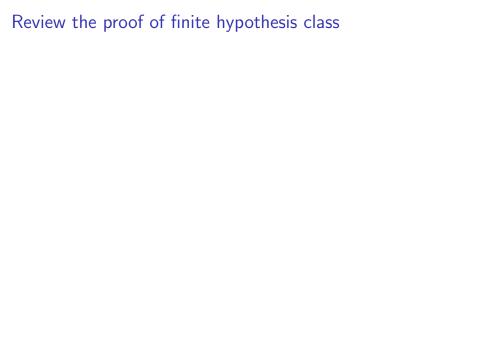
## Statistics Learning Theorem

2019.02.02



## Solution for problem

Let  $x_i$  be the Bernoulli variable with parameter  $p = Pr_{x \sim \mathcal{D}}\{x | h(x) \neq f(x)\}$ . That is,

$$x_i = \begin{cases} 1 & p \\ 0 & 1 - p \end{cases}$$

Then  $L_S(h) = \frac{\sum_{i=1}^m x_i}{m}$ . The iid assumption implies that

$$E_{S \sim \mathcal{D}^m}[L_S(h)] = E_{S \sim \mathcal{D}^m}\left[\frac{\sum_{i=1}^m x_i}{m}\right] = \frac{\sum_{i=1}^m E_{x_i}[x_i]}{m} = E[x_i] = p = L_{\mathcal{D},f}(h)$$

The  $E_{S \sim \mathcal{D}^m}[\frac{\sum_{i=1}^m x_i}{m}] = \frac{\sum_{i=1}^m E_{x_i}[x_i]}{m}$  follows from Fubini's theorem.

## PAC Learnability

A hypothesis class  $\mathcal H$  is PAC learnable if there exist a function  $m_{\mathcal H}:(0,1)^2\to N$  and a learning algorithm with the following property: For every  $,\delta\in(0,1)$ , for every distribution  $\mathcal D$  over  $\mathcal X$ , and for every labeling function