If we substitute $f - \mathbb{E}f$ instead of f, the result of Lecture 37 becomes:

$$\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n (f(x_i)-\mathbb{E}f)\right| \le \mathbb{E}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n (f(x_i)-\mathbb{E}f)\right| + \sqrt{\left(4(b-a)\mathbb{E}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n (f(x_i)-\mathbb{E}f)\right| + 2n\sigma^2\right)t} + (b-a)\frac{t}{3}$$

with probability at least $\geq 1 - e^{-t}$. Here, $a \le f \le b$ for all $f \in \mathcal{F}$ and $\sigma^2 = \sup_{f\in\mathcal{F}} \mathrm{Var}(f)$.

Now divide by n to get

$$\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n f(x_i)-\mathbb{E}f\right| \le \mathbb{E}\sup_{f\in\mathcal{F}}|\ldots| + \sqrt{\left(4(b-a)\mathbb{E}\sup_{f\in\mathcal{F}}|\ldots| + 2\sigma^2\right)\frac{t}{n}} + (b-a)\frac{t}{3n}$$

Compare this result to the Martingale-difference method (McDiarmid):

$$\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n f(x_i)-\mathbb{E}f\right| \le \mathbb{E}\sup_{f\in\mathcal{F}}|\ldots| + \sqrt{\frac{2(b-a)^2 t}{n}}$$

The term $2(b-a)^2$ is worse than $4(b-a)\mathbb{E}\sup_{f\in\mathcal{F}}|\ldots| + 2\sigma^2$.

An algorithm outputs $f_0 \in \mathcal{F}$, f_0 depends on data $x_1, \ldots, x_n$. What is $\mathbb{E}f_0$? Assume $0 \le f \le 1$ (loss function). Then

$$\left|\mathbb{E}f_0 - \frac{1}{n}\sum_{i=1}^n f_0(x_i)\right| \le \sup_{f\in\mathcal{F}}\left|\mathbb{E}f - \frac{1}{n}\sum_{i=1}^n f(x_i)\right| \le \text{ use Talagrand's inequality }.$$

What if we knew that $\mathbb{E}f_0 \le \varepsilon$ and the family $\mathcal{F}_\varepsilon = \{f \in \mathcal{F}, \mathbb{E}f \le \varepsilon\}$ is much smaller than $\mathcal{F}$. Then looking at $\sup_{f\in\mathcal{F}}\left|\mathbb{E}f - \frac{1}{n}\sum_{i=1}^n f(x_i)\right|$ is too conservative.

Pin down location of f_0. Pretend we know $\mathbb{E}f_0 \le \varepsilon$, $f_0 \in \mathcal{F}_\varepsilon$. Then with probability at least $1 - e^{-t}$,

$$\left|\mathbb{E}f_0 - \frac{1}{n}\sum_{i=1}^n f_0(x_i)\right| \le \sup_{f\in\mathcal{F}_\varepsilon}\left|\mathbb{E}f - \frac{1}{n}\sum_{i=1}^n f(x_i)\right|$$

$$\le \mathbb{E}\sup_{f\in\mathcal{F}_\varepsilon}\left|\mathbb{E}f - \frac{1}{n}\sum_{i=1}^n f(x_i)\right| + \sqrt{\left(4\mathbb{E}\sup_{f\in\mathcal{F}_\varepsilon}|\ldots| + 2\sigma_\varepsilon^2\right)\frac{t}{n}} + \frac{t}{3n}$$

where $\sigma_\varepsilon^2 = \sup_{f\in\mathcal{F}_\varepsilon} \mathrm{Var}(f)$. Note that for $f \in \mathcal{F}_\varepsilon$

$$\mathrm{Var}(f) = \mathbb{E}f^2 - (\mathbb{E}f)^2 \le \mathbb{E}f^2 \le \mathbb{E}f \le \varepsilon$$

since $0 \le f \le 1$.

Denote $\varphi(\varepsilon) = \mathbb{E}\sup_{f\in\mathcal{F}_\varepsilon}\left|\mathbb{E}f - \frac{1}{n}\sum_{i=1}^n f(x_i)\right|$. Then

$$\left|\mathbb{E}f_0 - \frac{1}{n}\sum_{i=1}^n f_0(x_i)\right| \le \varphi(\varepsilon) + \sqrt{(4\varphi(\varepsilon)+2\varepsilon)\frac{t}{n}} + \frac{t}{3n}$$

with probability at least $1 - e^{-t}$.

Take $\varepsilon = 2^{-k}$, $k = 0, 1, 2, \ldots$. Change $t \to t + 2\log(k+2)$. Then, for a fixed k, with probability at least $1 - e^{-t}\frac{1}{(k+2)^2}$,

$$\left|\mathbb{E}f_0 - \frac{1}{n}\sum_{i=1}^n f_0(x_i)\right| \le \varphi(\varepsilon) + \sqrt{(4\varphi(\varepsilon) + 2\varepsilon)\,\frac{t + 2\log(k+2)}{n}} + \frac{t + 2\log(k+2)}{3n}$$

For all $k \ge 0$, the statement holds with probability at least

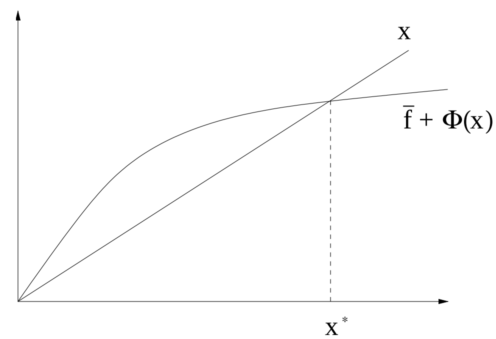
$$1 - \underbrace{\sum_{k=1}^{\infty}\frac{1}{(k+2)^2}}_{\frac{\pi^2}{6}-1} e^{-t} \ge 1 - e^{-t}$$

For f_0, find k such that $2^{-k-1} \le \mathbb{E}f_0 < 2^{-k}$ (hence, $2^{-k} \le 2\mathbb{E}f_0$). Use the statement for $\varepsilon_k = 2^{-k}$, $k \le \log_2\frac{1}{\mathbb{E}f_0}$.

$$\left|\mathbb{E}f_0 - \frac{1}{n}\sum_{i=1}^n f_0(x_i)\right| \le \varphi(\varepsilon_k) + \sqrt{(4\varphi(\varepsilon_k) + 2\varepsilon_k)\,\frac{t + 2\log(k+2)}{n}} + \frac{t + 2\log(k+2)}{3n}$$

$$\le \varphi(2\mathbb{E}f_0) + \sqrt{(4\varphi(2\mathbb{E}f_0) + 4\mathbb{E}f_0)\,\frac{t + 2\log(\log_2\frac{1}{\mathbb{E}f_0} + 2)}{n}} + \frac{t + 2\log(\log_2\frac{1}{\mathbb{E}f_0} + 2)}{3n} = \Phi(\mathbb{E}f_0)$$

Hence, $\mathbb{E}f_0 \le \frac{1}{n}\sum_{i=1}^n f_0(x_i) + \Phi(\mathbb{E}f_0)$. Denote $x = \mathbb{E}f_0$. Then $x \le \bar{f} + \Phi(x)$.



Theorem 38.1. *Let $0 \le f \le 1$ for all $f \in \mathcal{F}$. Define $\mathcal{F}_\varepsilon = \{f \in \mathcal{F}, \mathbb{E}f \le \varepsilon\}$ and $\varphi(\varepsilon) = \mathbb{E}\sup_{f\in\mathcal{F}_\varepsilon}\left|\mathbb{E}f - \frac{1}{n}\sum_{i=1}^n f(x_i)\right|$. Then, with probability at least $1 - e^{-t}$, for any $f_0 \in \mathcal{F}$, $\mathbb{E}f_0 \le x^*$, where x^* is the largest solution of*

$$x^* = \frac{1}{n}\sum_{i=1}^n f_0(x_i) + \Phi(x^*).$$

Main work is to find $\varphi(\varepsilon)$. Consider the following example.

Example 38.1. If

$$\sup_{x_1,\ldots,x_n}\log\mathcal{D}(\mathcal{F}, u, d_x) \le \mathcal{D}(\mathcal{F}, u),$$

then

$$\mathbb{E}\sup_{f\in\mathcal{F}_\varepsilon}\left|\mathbb{E}f - \frac{1}{n}\sum_{i=1}^n f(x_i)\right| \le \frac{k}{\sqrt{n}}\int_0^{\sqrt{\varepsilon}}\log^{1/2}\mathcal{D}(\mathcal{F},\varepsilon)d\varepsilon.$$