Lemma 15.1. *Let ξ, ν - random variables. Assume that*

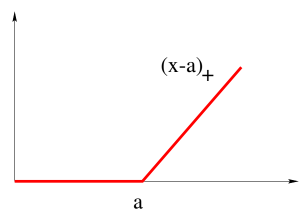
$$\mathbb{P}(\nu \geq t) \leq \Gamma e^{-\gamma t}$$

where $\Gamma \geq 1$, $t \geq 0$, and $\gamma > 0$. Furthermore, for all $a > 0$ assume that

$$\mathbb{E}\phi(\xi) \leq \mathbb{E}\phi(\nu)$$

where $\phi(x) = (x-a)_+$. Then

$$\mathbb{P}(\xi \geq t) \leq \Gamma \cdot e \cdot e^{-\gamma t}.$$



Proof. Since $\phi(x) = (x-a)_+$, we have $\phi(\xi) \geq \phi(t)$ whenever $\xi \geq t$.

$$\begin{aligned}
\mathbb{P}(\xi \geq t) &\leq \mathbb{P}(\phi(\xi) \geq \phi(t)) \\
&\leq \frac{\mathbb{E}\phi(\xi)}{\phi(t)} \leq \frac{\mathbb{E}\phi(\nu)}{\phi(t)} = \frac{\mathbb{E}(\nu-a)_+}{(t-a)_+}
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathbb{E}(\nu-a)_+ &= \mathbb{E}\int_0^{(\nu-a)_+} 1dx \\
&= \mathbb{E}\int_0^{\infty} I(x \leq (\nu-a)_+)dx \\
&= \int_0^{\infty} \mathbb{E}I(x \leq (\nu-a)_+)dx \\
&= \int_0^{\infty} \mathbb{P}((\nu-a)_+ \geq x)\, dx \\
&= \int_0^{\infty} \mathbb{P}(\nu \geq a+x)\, dx \\
&\leq \int_0^{\infty} \Gamma e^{-\gamma a-\gamma x} dx = \frac{\Gamma e^{-\gamma a}}{\gamma}.
\end{aligned}$$

Hence,

$$\mathbb{P}(\xi \geq t) \leq \frac{\Gamma e^{-\gamma a}}{\gamma(t-a)_+} = \frac{\Gamma \cdot e \cdot e^{-\gamma t}}{1} = \Gamma \cdot e \cdot e^{-\gamma t}$$

where we chose optimal $a = t - \frac{1}{\gamma}$ to minimize $\frac{\Gamma e^{-\gamma a}}{\gamma}$. $\square$

Lemma 15.2. *Let $x = (x_1, \ldots, x_n)$, $x' = (x'_1, \ldots, x'_n)$. If for functions $\varphi_1(x, x')$, $\varphi_2(x, x')$, $\varphi_3(x, x')$*

$$\mathbb{P}\left(\varphi_1(x, x') \geq \varphi_2(x, x') + \sqrt{\varphi_3(x, x') \cdot t}\right) \leq \Gamma e^{-\gamma t}$$

then

$$\mathbb{P}\left(\mathbb{E}_{x'}\varphi_1(x, x') \geq \mathbb{E}_{x'}\varphi_2(x, x') + \sqrt{\mathbb{E}_{x'}\varphi_3(x, x') \cdot t}\right) \leq \Gamma \cdot e \cdot e^{-\gamma t}.$$

(i.e. if the inequality holds, then it holds with averaging over one of the copies)

Proof. First, note that $\sqrt{ab} = \inf_{\delta>0}(\delta a + \frac{b}{4\delta})$ with $\delta_* = \sqrt{\frac{b}{4a}}$ achieving the infima. Hence,

$$\begin{aligned}
\{\varphi_1 \geq \varphi_2 + \sqrt{\varphi_3 t}\} &= \{\exists \delta > 0, \varphi_1 \geq \varphi_2 + \delta\varphi_3 + \frac{t}{4\delta}\} \\
&= \{\exists \delta > 0, (\varphi_1 - \varphi_2 - \delta\varphi_3)4\delta \geq t\} \\
&= \{\underbrace{\sup_{\delta>0}(\varphi_1 - \varphi_2 - \delta\varphi_3)4\delta}_{\nu} \geq t\}
\end{aligned}$$

and similarly

$$\{\mathbb{E}_{x'}\varphi_1 \geq \mathbb{E}_{x'}\varphi_2 + \sqrt{\mathbb{E}_{x'}\varphi_3 t}\} = \{\underbrace{\sup_{\delta>0}(\mathbb{E}_{x'}\varphi_1 - \mathbb{E}_{x'}\varphi_2 - \delta\mathbb{E}_{x'}\varphi_3)4\delta}_{\xi} \geq t\}.$$

By assumption, $\mathbb{P}(\nu \geq t) \leq \Gamma e^{-\gamma t}$. We want to prove $\mathbb{P}(\xi \geq t) \leq \Gamma \cdot e \cdot e^{-\gamma t}$. By the previous lemma, we only need to check whether $\mathbb{E}\phi(\xi) \leq \mathbb{E}\phi(\nu)$.

$$\begin{aligned}
\xi &= \sup_{\delta>0} \mathbb{E}_{x'}(\varphi_1 - \varphi_2 - \delta\varphi_3)4\delta \\
&\leq \mathbb{E}_{x'} \sup_{\delta>0}(\varphi_1 - \varphi_2 - \delta\varphi_3)4\delta \\
&= \mathbb{E}_{x'}\nu
\end{aligned}$$

Thus,

$$\phi(\xi) \leq \phi(\mathbb{E}_{x'}\nu) \leq \mathbb{E}_{x'}\phi(\nu)$$

by Jensen's inequality (ϕ is convex). Hence,

$$\mathbb{E}\phi(\xi) \leq \mathbb{E}\mathbb{E}_{x'}\phi(\nu) = \mathbb{E}\phi(\nu).$$

□

We will now use Lemma 15.2. Let $\mathcal{F} = \{f : \mathcal{X} \mapsto [c, c+1]\}$. Let $x_1, \ldots, x_n, x'_1, \ldots, x'_n$ be i.i.d. random variables. Define

$$F = \{(f(x_1) - f(x'_1), \ldots, f(x_n) - f(x'_n)) : f \in \mathcal{F}\} \subseteq [-1, 1]^n.$$

Define

$$d(f,g) = \left(\frac{1}{n}\sum_{i=1}^{n}\left((f(x_i) - f(x_i')) - (g(x_i) - g(x_i'))\right)^2\right)^{1/2}.$$

In Lecture 14, we proved

$$\mathbb{P}_\varepsilon\Bigg(\forall f \in \mathcal{F}, \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(f(x_i) - f(x_i')) \le \frac{2^{9/2}}{\sqrt{n}}\int_0^{d(0,f)} \log^{1/2}\mathcal{D}(\mathcal{F},\varepsilon,d)d\varepsilon$$
$$+2^{7/2}d(0,f)\sqrt{\frac{t}{n}}\Bigg) \ge 1 - e^{-t}.$$

Complement of the above is

$$\mathbb{P}_\varepsilon\left(\exists f \in \mathcal{F}, \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(f(x_i) - f(x_i')) \ge \frac{2^{9/2}}{\sqrt{n}}\int_0^{d(0,f)} \log^{1/2}\mathcal{D}(\mathcal{F},\varepsilon,d)d\varepsilon + 2^{7/2}d(0,f)\sqrt{\frac{t}{n}}\right) \le e^{-t}.$$

Taking expectation with respect to x, x', we get

$$\mathbb{P}\left(\exists f \in \mathcal{F},\ \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(f(x_i) - f(x_i')) \ge \frac{2^{9/2}}{\sqrt{n}}\int_0^{d(0,f)} \log^{1/2}\mathcal{D}(\mathcal{F},\varepsilon,d)d\varepsilon + 2^{7/2}d(0,f)\sqrt{\frac{t}{n}}\right) \le e^{-t}.$$

Hence (see below)

$$\mathbb{P}\left(\exists f \in \mathcal{F},\ \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - f(x_i')) \ge \frac{2^{9/2}}{\sqrt{n}}\int_0^{d(0,f)} \log^{1/2}\mathcal{D}(\mathcal{F},\varepsilon,d)d\varepsilon + 2^{7/2}d(0,f)\sqrt{\frac{t}{n}}\right) \le e^{-t}.$$

To see why the above step holds, notice that $d(f,g)$ is invariant under permutations $x_i \leftrightarrow x_i'$. We can remove ε_i since x and x' are i.i.d and we can switch x_i and x_i'. To the right of "$\ge$" sign, only distance $d(f,g)$ depends on x, x', but it's invariant to the permutations.

By Lemma 15.2 (minus technical detail "$\exists f$"),

$$\mathbb{P}\Bigg(\exists f \in \mathcal{F},\ \mathbb{E}_{x'}\frac{1}{n}\sum_{i=1}^{n}(f(x_i) - f(x_i')) \ge \mathbb{E}_{x'}\frac{2^{9/2}}{\sqrt{n}}\int_0^{d(0,f)} \log^{1/2}\mathcal{D}(\mathcal{F},\varepsilon,d)d\varepsilon$$
$$+2^{7/2}\sqrt{\frac{\mathbb{E}_{x'}d(0,f)^2 t}{n}}\Bigg) \le e \cdot e^{-t},$$

where

$$\mathbb{E}_{x'}\frac{1}{n}\sum_{i=1}^{n}(f(x_i) - f(x_i')) = \frac{1}{n}\sum_{i=1}^{n}f(x_i) - \mathbb{E}f$$

and

$$\mathbb{E}_{x'}d(0,f)^2 = \mathbb{E}_{x'}\frac{1}{n}\sum_{i=1}^{n}(f(x_i) - f(x_i'))^2.$$

The Dudley integral above will be bounded by something non-random in the later lectures.