Assume we have space $\mathcal{X}$ and a class of functions $\mathcal{F} = \{f : \mathcal{X} \mapsto \mathbb{R}\}$, not necessarily bounded. Define

$$Z(x) = Z(x_1, \ldots, x_n) = \sup_{f \in \mathcal{F}} \sum f(x_i)$$

(or $\sup_{f\in\mathcal{F}} |\sum f(x_i)|$).

Example 1. $f \to \frac{1}{n}(f - \mathbb{E}f)$. $Z(x) = \sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n f(x_i) - \mathbb{E}f$.

Consider $x' = (x'_1, \ldots, x'_n)$, an independent copy of x. Let

$$V(x) = \mathbb{E}_{x'} \sup_{f\in\mathcal{F}} \sum_{i=1}^n (f(x_i) - f(x'_i))^2$$

be "random uniform variance" (unofficial name)

Theorem 28.1.

$$\mathbb{P}\left(Z(x) \geq \mathbb{E}Z(x) + 2\sqrt{V(x)t}\right) \leq 4e \cdot e^{-t/4}$$
$$\mathbb{P}\left(Z(x) \leq \mathbb{E}Z(x) - 2\sqrt{V(x)t}\right) \leq 4e \cdot e^{-t/4}$$

Recall the Symmetrization lemma:

Lemma 28.1. $\xi_1, \xi_2, \xi_3(x, x') : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, $\xi'_i = \mathbb{E}_{x'}\xi_i$. *If*

$$\mathbb{P}\left(\xi_1 \geq \xi_2 + \sqrt{\xi_3 t}\right) \leq \Gamma e^{-\gamma t},$$

then

$$\mathbb{P}\left(\xi'_1 \geq \xi'_2 + \sqrt{\xi'_3 t}\right) \leq \Gamma e \cdot e^{-\gamma t}.$$

We have

$$\mathbb{E}Z(x) = \mathbb{E}_{x'}Z(x') = \mathbb{E}_{x'} \sup_{f\in\mathcal{F}} \sum_{i=1}^n f(x'_i)$$

and

$$V(x) = \mathbb{E}_{x'} \sup_{f\in\mathcal{F}} \sum_{i=1}^n (f(x_i) - f(x'_i))^2.$$

Use the Symmetrization Lemma with $\xi_1 = Z(x)$, $\xi_2 = Z(x')$, and

$$\xi_3 = \sup_{f\in\mathcal{F}} \sum_{i=1}^n (f(x_i) - f(x'_i))^2.$$

It is enough to prove that

$$\mathbb{P}\left(Z(x) \geq Z(x') + 2\sqrt{t \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} (f(x_i) - f(x_i'))^2}\right) \leq 4e^{-t/4},$$

i.e.

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}} \sum_{i=1}^{n} f(x_i) \geq \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} f(x_i') + 2\sqrt{t \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} (f(x_i) - f(x_i'))^2}\right) \leq 4e^{-t/4}.$$

If we switch $x_i \leftrightarrow x_i'$, nothing changes, so we can switch randomly. Implement the permutation $x_i \leftrightarrow x_i'$:

$$I = f(x_i') + \varepsilon_i(f(x_i) - f(x_i'))$$

$$II = f(x_i) - \varepsilon_i(f(x_i) - f(x_i'))$$

where $\varepsilon_i = 0, 1$. Hence,

(1) If $\varepsilon_i = 1$, then $I = f(x_i)$ and $II = f(x_i')$.

(2) If $\varepsilon_i = 0$, then $I = f(x_i')$ and $II = f(x_i)$.

Take $\varepsilon_1 \ldots \varepsilon_n$ i.i.d. with $\mathbb{P}(\varepsilon_i = 0) = \mathbb{P}(\varepsilon_i = 1) = 1/2$.

$$\begin{aligned}
&\mathbb{P}_{x,x'}\left(\sup_{f\in\mathcal{F}} \sum_{i=1}^{n} f(x_i) \geq \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} f(x_i') + 2\sqrt{t \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} (f(x_i) - f(x_i'))^2}\right) \\
&= \mathbb{P}_{x,x',\varepsilon}\left(\sup_{f\in\mathcal{F}} \sum_{i=1}^{n} (f(x_i') + \varepsilon_i(f(x_i) - f(x_i'))) \geq \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} (f(x_i) - \varepsilon_i(f(x_i) - f(x_i')))\right. \\
&\qquad\qquad \left. + 2\sqrt{t \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} (f(x_i) - f(x_i'))^2}\right) \\
&= \mathbb{E}_{x,x'}\mathbb{P}_{\varepsilon}\left(\sup_{f\in\mathcal{F}} \ldots \geq \sup_{f\in\mathcal{F}} \ldots + 2\sqrt{\ldots} \text{ for fixed } x, x'\right)
\end{aligned}$$

Define

$$\Phi_1(\varepsilon) = \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} (f(x_i') + \varepsilon_i(f(x_i) - f(x_i')))$$

and

$$\Phi_2(\varepsilon) = \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} (f(x_i) - \varepsilon_i(f(x_i) - f(x_i'))).$$

$\Phi_1(\varepsilon), \Phi_2(\varepsilon)$ are convex and Lipschitz with $L = \sup_{f\in\mathcal{F}} \sqrt{\sum_{i=1}^n (f(x_i) - f(x_i'))^2}$. Moreover, $Median(\Phi_1) = Median(\Phi_2)$ and $\Phi_1(\varepsilon_1, \ldots, \varepsilon_n) = \Phi_2(1-\varepsilon_1, \ldots, 1-\varepsilon_n)$. Hence,

$$\mathbb{P}_\varepsilon\left(\Phi_1 \le M(\Phi_1) + L\sqrt{t}\right) \ge 1 - 2e^{-t/4}$$

and

$$\mathbb{P}_\varepsilon\left(\Phi_2 \le M(\Phi_2) - L\sqrt{t}\right) \ge 1 - 2e^{-t/4}.$$

With probability at least $1 - 4e^{-t/4}$ both above inequalities hold:

$$\Phi_1 \le M(\Phi_1) + L\sqrt{t} = M(\Phi_2) + L\sqrt{t} \le \Phi_2 + 2L\sqrt{t}.$$

Thus,

$$\mathbb{P}_\varepsilon\left(\Phi_1 \ge \Phi_2 + 2L\sqrt{t}\right) \le 4e^{-t/4}$$

and

$$\mathbb{P}_{x,x',\varepsilon}\left(\Phi_1 \ge \Phi_2 + 2L\sqrt{t}\right) \le 4e^{-t/4}.$$

The "random uniform variance" is

$$V(x) = \mathbb{E}_{x'} \sup_{f\in\mathcal{F}} \sum_{i=1}^n (f(x_i) - f(x_i'))^2.$$

For example, if $\mathcal{F} = \{f\}$, then

$$\begin{aligned}
\frac{1}{n}V(x) &= \frac{1}{n}\mathbb{E}_{x'} \sum_{i=1}^n (f(x_i) - f(x_i'))^2 \\
&\frac{1}{n}\sum_{i=1}^n \left(f(x_i)^2 - 2f(x_i)\mathbb{E}f + \mathbb{E}f^2\right) \\
&= \bar{f^2} - 2\bar{f}\mathbb{E}f + \mathbb{E}f^2 \\
&= \underbrace{\bar{f^2} - (\bar{f})^2}_{\text{sample variance}} + \underbrace{(\bar{f})^2 - 2\bar{f}\mathbb{E}f + (\mathbb{E}f)^2}_{(\bar{f}-\mathbb{E}f)^2} + \underbrace{\mathbb{E}f^2 - (\mathbb{E}f)^2}_{\text{variance}}
\end{aligned}$$