Lemma 30.1. *Let*

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

and $a \le f \le b$ *for all* $f \in \mathcal{F}$. *Then*

$$\mathbb{P}\left(V \le 4\mathbb{E}V + (b-a)^2 t\right) \ge 1 - 4\cdot 2^{-t}.$$

Proof. Consider M-median of V, i.e. $\mathbb{P}(V \ge M) \ge 1/2$, $\mathbb{P}(V \le M) \ge 1/2$. Let $A = \{y \in \mathcal{X}^n, V(y) \le M\} \subseteq \mathcal{X}^n$. Hence, A consists of points with typical behavior. We will use control by 2 points to show that any other point is close to these two points.

By control by 2 points,

$$\mathbb{P}(d(A,A,x) \ge t) \le \frac{1}{\mathbb{P}(A)\,\mathbb{P}(A)} \cdot 2^{-t} \le 4\cdot 2^{-t}$$

Take any $x \in \mathcal{X}^n$. With probability at least $1 - 4\cdot 2^{-t}$, $d(A,A,x) \le t$. Hence, we can find $y^1 \in A, y^2 \in A$ such that $\mathrm{card}\,\{i \le n, x_i \ne y_i^1, x_i \ne y_i^2\} \le t$.

Let

$$I_1 = \{i \le n : x_i = y_i^1\},\ \ I_2 = \{i \le n : x_i \ne y_i^1, x_i = y_i^2\},$$

and

$$I_3 = \{i \le n : x_i \ne y_i^1, x_i \ne y_i^2\}$$

Then we can decompose V as follows

$$\begin{aligned}
V(x) &= \mathbb{E}_{x'} \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} (f(x_i) - f(x_i'))^2 \\
&= \mathbb{E}_{x'} \sup_{f\in\mathcal{F}} \left[\sum_{i\in I_1} (f(x_i) - f(x_i'))^2 + \sum_{i\in I_2} (f(x_i) - f(x_i'))^2 + \sum_{i\in I_3} (f(x_i) - f(x_i'))^2\right] \\
&\le \mathbb{E}_{x'} \sup_{f\in\mathcal{F}} \sum_{i\in I_1} (f(x_i) - f(x_i'))^2 + \mathbb{E}_{x'} \sup_{f\in\mathcal{F}} \sum_{i\in I_2} (f(x_i) - f(x_i'))^2 + \mathbb{E}_{x'} \sup_{f\in\mathcal{F}} \sum_{i\in I_3} (f(x_i) - f(x_i'))^2 \\
&\le \mathbb{E}_{x'} \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} (f(y_i^1) - f(x_i'))^2 + \mathbb{E}_{x'} \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} (f(y_i^2) - f(x_i'))^2 + (b-a)^2 t \\
&= V(y^1) + V(y^2) + (b-a)^2 t \\
&\le M + M + (b-a)^2 t
\end{aligned}$$

because $y^1, y^2 \in A$. Hence,

$$\mathbb{P}\left(V(x) \leq 2M + (b-a)^2 t\right) \geq 1 - 4 \cdot 2^{-t}.$$

Finally, $M \leq 2\mathbb{E}V$ because

$$\mathbb{P}(V \geq 2\mathbb{E}V) \leq \frac{\mathbb{E}V}{2\mathbb{E}V} = \frac{1}{2} \qquad \text{while} \qquad \mathbb{P}(V \geq M) \geq \frac{1}{2}.$$

□

Now, let $Z(x) = \sup_{f \in \mathcal{F}} |\sum_{i=1}^n f(x_i)|$. Then

$$Z(x) \underbrace{\leq}_{\text{with prob. } \geq 1-(4e)e^{-t/4}} \mathbb{E}Z + 2\sqrt{V(x)t} \underbrace{\leq}_{\text{with prob. } \geq 1-4\cdot 2^{-t}} \mathbb{E}Z + 2\sqrt{(4\mathbb{E}V + (b-a)^2 t)t}.$$

Using inequality $\sqrt{c+d} \leq \sqrt{c} + \sqrt{d}$,

$$Z(x) \leq \mathbb{E}Z + 4\sqrt{\mathbb{E}Vt} + 2(b-a)t$$

with high probability.

We proved Talagrand's concentration inequality for empirical processes:

Theorem 30.1. *Assume $a \leq f \leq b$ for all $f \in \mathcal{F}$. Let $Z = \sup_{f \in \mathcal{F}} |\sum_{i=1}^n f(x_i)|$ and $V = \sup_{f \in \mathcal{F}} \sum_{i=1}^n (f(x_i) - f(x_i'))^2$. Then*

$$\mathbb{P}\left(Z \leq \mathbb{E}Z + 4\sqrt{\mathbb{E}Vt} + 2(b-a)t\right) \geq 1 - (4e)e^{-t/4} - 4 \cdot 2^{-t}.$$

This is an analog of Bernstein's inequality:

$$4\sqrt{\mathbb{E}Vt} \longrightarrow \text{Gaussian behavior}$$

$$2(b-a)t \longrightarrow \text{Poisson behavior}$$

Now, consider the following lower bound on V.

$$\begin{aligned}
V &= \mathbb{E}\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}(f(x_i)-f(x_i'))^2 \\
&> \sup_{f\in\mathcal{F}}\mathbb{E}\sum_{i=1}^{n}(f(x_i)-f(x_i'))^2 \\
&= \sup_{f\in\mathcal{F}} n\mathbb{E}(f(x_1)-f(x_1'))^2 \\
&= \sup_{f\in\mathcal{F}} 2n\mathrm{Var}(f) = 2n\sup_{f\in\mathcal{F}}\mathrm{Var}(f) = 2n\sigma^2
\end{aligned}$$

As for the upper bound,

$$\begin{aligned}
\mathbb{E}\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}(f(x_i)-f(x_i'))^2 &= \mathbb{E}\sup_{f\in\mathcal{F}}\left(\sum_{i=1}^{n}(f(x_i)-f(x_i'))^2 - 2n\mathrm{Var}(f) + 2n\mathrm{Var}(f)\right) \\
&\le \mathbb{E}\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\left[(f(x_i)-f(x_i'))^2 - \mathbb{E}(f(x_i)-f(x_i'))^2\right] + 2n\sup_{f\in\mathcal{F}}\mathrm{Var}(f) \\
&\text{(by symmetrization)} \\
&\le 2\mathbb{E}\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\varepsilon_i(f(x_i)-f(x_i'))^2 + 2n\sigma^2 \\
&\le 2\mathbb{E}\left(\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\varepsilon_i(f(x_i)-f(x_i'))^2\right)_+ + 2n\sigma^2
\end{aligned}$$

Note that the square function $[-(b-a),(b-a)]\mapsto\mathbb{R}$ is a contraction. Its largest derivative on $[-(b-a),(b-a)]$ is at most $2(b-a)$. Note that $|f(x_i)-f(x_i')|\le b-a$. Hence,

$$\begin{aligned}
2\mathbb{E}\left(\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\varepsilon_i(f(x_i)-f(x_i'))^2\right)_+ + 2n\sigma^2 &\le 2\cdot 2(b-a)\mathbb{E}\left(\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\varepsilon_i(f(x_i)-f(x_i'))\right)_+ + 2n\sigma^2 \\
&\le 4(b-a)\mathbb{E}\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\varepsilon_i|f(x_i)-f(x_i')| + 2n\sigma^2 \\
&\le 4(b-a)\cdot 2\mathbb{E}\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}\varepsilon_i|f(x_i)| + 2n\sigma^2 \\
&= 8(b-a)\mathbb{E}Z + 2n\sigma^2
\end{aligned}$$

We have proved the following

Lemma 30.2.

$$\mathbb{E}V \leq 8(b-a)\mathbb{E}Z + 2n\sigma^2,$$

where $\sigma^2 = \sup_{f\in\mathcal{F}} \mathrm{Var}(f)$.

Corollary 30.1. *Assume* $a \leq f \leq b$ *for all* $f \in \mathcal{F}$. *Let* $Z = \sup_{f\in\mathcal{F}} |\sum_{i=1}^n f(x_i)|$ *and* $\sigma^2 = \sup_{f\in\mathcal{F}} \mathrm{Var}(f)$. *Then*

$$\mathbb{P}\left(Z \leq \mathbb{E}Z + 4\sqrt{(8(b-a)\mathbb{E}Z + 2n\sigma^2)t} + 2(b-a)t\right) \geq 1 - (4e)e^{-t/4} - 4\cdot 2^{-t}.$$

Using other approaches, one can get better constants:

$$\mathbb{P}\left(Z \leq \mathbb{E}Z + \sqrt{(4(b-a)\mathbb{E}Z + 2n\sigma^2)t} + (b-a)\frac{t}{3}\right) \geq 1 - e^{-t}.$$