Recall the tensorization of entropy lemma we proved previously. Let $x_1, \cdots, x_n$ be independent random variables, $x'_1, \cdots, x'_n$ be their independent copies, $Z = Z(x_1, \cdots, x_n)$, $Z^i = (x_1, \cdots, x_{i-1}, x'_i, x_{i+1}, \cdots, x_n)$, and $\phi(x) = e^x - x - 1$. We have $\mathbb{E}e^{\lambda Z} - \mathbb{E}e^{\lambda Z} \log \mathbb{E}e^{\lambda Z} \le \mathbb{E}e^{\lambda Z} \sum_{i=1}^n \phi(-\lambda(Z - Z^i))$. We will use the tensorization of entropy technique to prove the following Hoeffding-type inequality. This theorem is Theorem 9 of `Pascal Massart. About the constants in Talagrand's concentration inequalities for empirical processes. The Annals of Probability, 2000, Vol 28, No. 2, 863-884`.

Theorem 41.1. *Let $\mathcal{F}$ be a finite set of functions $|\mathcal{F}| < \infty$. For any $f = (f_1, \cdots, f_n) \in \mathcal{F}$, $a_i \le f_i \le b_i$, $L = \sup_f \sum_{i=1}^n (b_i - a_i)^2$, and $Z = \sup_f \sum_{i=1}^n f_i$. Then $\mathbb{P}(Z \ge \mathbb{E}Z + \sqrt{2Lt}) \le e^{-t}$.*

Proof. Let

$$
\begin{aligned}
Z^i &= \sup_{f \in \mathcal{F}} \left(a_i + \sum_{j \ne i} f_j \right) \\
Z &= \sup_{f \in \mathcal{F}} \sum_{i=1}^n f_i \stackrel{\text{def.}}{=} \sum_{i=1}^n f_i^\circ.
\end{aligned}
$$

It follows that

$$0 \le Z - Z^i \le \sum_i f_i^\circ - \sum_{j \ne i} f_j^\circ - a_i = f_i^\circ - a_i \le b_i(f^\circ) - a_i(f^\circ).$$

Since $\frac{\phi(x)}{x^2} = \frac{e^x - x - 1}{x^2}$ is increasing in $\mathbb{R}$ and $\lim_{x \to 0} \frac{\phi(x)}{x^2} \to \frac{1}{2}$, it follows that $\forall x < 0$, $\phi(x) \le \frac{1}{2}x^2$, and

$$
\begin{aligned}
\mathbb{E}e^{\lambda Z}\lambda Z - \mathbb{E}e^{\lambda Z} \log \mathbb{E}e^{\lambda Z} &\le \mathbb{E}e^{\lambda Z} \sum_i \phi\left(-\lambda(Z - Z^i)\right) \\
&\le \frac{1}{2}\mathbb{E}e^{\lambda Z} \sum_i \lambda^2 (Z - Z^i)^2 \\
&\le \frac{1}{2} L \lambda^2 \mathbb{E}e^{\lambda Z}.
\end{aligned}
$$

Center Z, and we get

$$\mathbb{E}e^{\lambda(Z - \mathbb{E}Z)}\lambda(Z - \mathbb{E}Z) - \log \mathbb{E}e^{\lambda(Z - \mathbb{E}Z)} \le \frac{1}{2} L \lambda^2 \mathbb{E}e^{\lambda(Z - \mathbb{E}Z)}.$$

Let $F(\lambda) = \mathbb{E}e^{\lambda(Z-\mathbb{E}Z)}$. It follows that $F'_\lambda(\lambda) = \mathbb{E}e^{\lambda(Z-\mathbb{E}Z)}(Z - \mathbb{E}Z)$ and

$$\begin{aligned}
\lambda F'_\lambda(\lambda) - F(\lambda)\log F(\lambda) &\leq \frac{1}{2}L\lambda^2 F(\lambda) \\
\frac{1}{\lambda}\frac{F'_\lambda(\lambda)}{F(\lambda)} - \frac{1}{\lambda^2}\log F(\lambda) &\leq \frac{1}{2}L \\
\left(\frac{1}{\lambda}\log F(\lambda)\right)'_\lambda &\leq \frac{1}{2}L \\
\frac{1}{\lambda}\log F(\lambda) &= \frac{1}{t}\log F(t)\Big|_{t\to 0} + \int_0^\lambda \left(\frac{1}{t}\log F(t)\right)'_t dt \\
&\leq \frac{1}{2}L\lambda \\
F(\lambda) &\leq \exp(\frac{1}{2}L\lambda^2).
\end{aligned}$$

By Chebychev inequality, and minimize over λ, we get

$$\begin{aligned}
\mathbb{P}(Z \geq \mathbb{E}Z + t) &\leq e^{-\lambda t}\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)} \\
&\leq e^{-\lambda t}e^{\frac{1}{2}L\lambda^2} \\
\overset{\text{minimize over } \lambda}{\mathbb{P}(Z \geq \mathbb{E}Z + t)} &\leq e^{-t^2/(2L)}
\end{aligned}$$

□

Let f_i above be Rademacher random variables and apply Hoeffding's inequality, we get $\mathbb{P}(Z \geq \mathbb{E}Z + \sqrt{Lt/2}) \leq e^{-t}$. As a result, The above inequality improves the constant of Hoeffding's inequality.

The following Bennett type concentration inequality is Theorem 10 of

`Pascal Massart. About the constants in Talagrand's concentration inequalities for empirical processes. The Annals of Probability, 2000, Vol 28, No. 2, 863-884.`

Theorem 41.2. *Let $\mathcal{F}$ be a finite set of functions $|\mathcal{F}| < \infty$. $\forall f = (f_1, \cdots, f_n) \in \mathcal{F}$, $0 \leq f_i \leq 1$, $Z = \sup_f \sum_{i=1}^n f_i$, and define h as $h(u) = (1+u)\log(1+u) - u$ where $u \geq 0$. Then $\mathbb{P}(Z \geq \mathbb{E}Z + x) \leq e^{-\mathbb{E}Z \cdot h(x/\mathbb{E}Z)}$.*

Proof. Let

$$\begin{aligned}
Z &= \sup_{f\in\mathcal{F}} \sum_{i=1}^n f_i \overset{\text{def.}}{=} \sum_{i=1}^n f_i^\circ \\
Z^i &= \sup_{f\in\mathcal{F}} \sum_{j\neq i} f_j.
\end{aligned}$$

It follows that $0 \leq Z - Z^i \leq f_i^\circ \leq 1$. Since $\phi = e^x - x - 1$ is a convex function of x,

$$\phi(-\lambda(Z - Z^i)) = \phi(-\lambda \cdot (Z - Z^i) + 0 \cdot (1 - (Z - Z^i))) \leq (Z - Z^i)\phi(-\lambda)$$

and

$$\begin{aligned}
\mathbb{E}\left(\lambda Z e^{\lambda Z}\right) - \mathbb{E}e^{\lambda Z}\log \mathbb{E}e^{\lambda Z} &\leq \mathbb{E}\left(e^{\lambda Z}\sum_{i=1}^{n}\phi\left(-\lambda(Z-Z^i)\right)\right) \\
&\leq \mathbb{E}\left(e^{\lambda Z}\phi(-\lambda)\sum_i\left(Z-Z^i\right)\right) \\
&\leq \phi(-\lambda)\mathbb{E}\left(e^{\lambda Z}\cdot\sum_i f_i^\circ\right) \\
&= \phi(-\lambda)\mathbb{E}\left(Z\cdot e^{\lambda Z}\right).
\end{aligned}$$

Set $\tilde{Z} = Z - \mathbb{E}Z$ (i.e., center Z), and we get

$$\begin{aligned}
\mathbb{E}\left(\lambda\tilde{Z}e^{\lambda\tilde{Z}}\right) - \mathbb{E}e^{\lambda\tilde{Z}}\log\mathbb{E}e^{\lambda\tilde{Z}} &\leq \phi(-\lambda)\mathbb{E}\left(\tilde{Z}\cdot e^{\lambda\tilde{Z}}\right) \leq \phi(-\lambda)\mathbb{E}\left(\left(\tilde{Z}+\mathbb{E}Z\right)\cdot e^{\lambda\tilde{Z}}\right) \\
\left(\lambda-\phi(-\lambda)\right)\mathbb{E}\left(\tilde{Z}e^{\lambda\tilde{Z}}\right) - \mathbb{E}e^{\lambda\tilde{Z}}\log\mathbb{E}e^{\lambda\tilde{Z}} &\leq \phi(-\lambda)\cdot\mathbb{E}Z\cdot\mathbb{E}e^{\lambda\tilde{Z}}.
\end{aligned}$$

Let $v = \mathbb{E}Z$, $F(\lambda) = \mathbb{E}e^{\lambda\tilde{Z}}$, $\Psi = \log F$, and we get

$$\begin{aligned}
\left(\lambda-\phi(-\lambda)\right)\frac{F'_\lambda(\lambda)}{F(\lambda)} - \log F(\lambda) &\leq v\phi(-\lambda) \\
\left(\lambda-\phi(-\lambda)\right)\left(\log F(\lambda)\right)'_\lambda - \log F(\lambda) &\leq v\phi(-\lambda). \qquad (41.1)
\end{aligned}$$

Solving the differential equation

$$\left(\lambda-\phi(-\lambda)\right)\left(\underbrace{\log F(\lambda)}_{\Psi_0}\right)'_\lambda - \underbrace{\log F(\lambda)}_{\Psi_0} = v\phi(-\lambda), \qquad (41.2)$$

yields $\Psi_0 = v\cdot\phi(\lambda)$. We will proceed to show that Ψ satisfying 41.1 has the property $\Psi \leq \Psi_0$:

$$\begin{aligned}
\overset{\text{Substract 41.2 from 41.1, and let } \Psi_1=\Psi-\Psi_0}{(1-e^{-\lambda})\Psi_1' - \Psi_1} &\leq 0 \\
\scriptstyle \left(e^\lambda-1\right)\left(1-e^{-\lambda}\right)\frac{1}{e^\lambda-1}=1-e^{-\lambda}\ \text{,and}\ \left(e^\lambda-1\right)\left(1-e^{-\lambda}\right)\frac{e^\lambda}{\left(e^\lambda-1\right)^2}=1 & \\
\left(e^\lambda-1\right)\left(1-e^{-\lambda}\right)\underbrace{\left(\frac{1}{e^\lambda-1}\Psi_1' - \frac{e^\lambda}{\left(e^\lambda-1\right)^2}\Psi_1\right)}_{\left(\frac{\Psi_1}{e^\lambda-1}\right)'_\lambda} &\leq 0 \\
\frac{\Psi_1(\lambda)}{e^\lambda-1} &\leq \lim_{\lambda\to 0}\frac{\Psi_1(\lambda)}{e^\lambda-1} = 0.
\end{aligned}$$

It follows that $\Psi \leq v\phi(\lambda)$, and $F = \mathbb{E}e^{\lambda Z} \leq e^{v\phi(\lambda)}$. By Chebychev's inequality, $\mathbb{P}\left(Z \geq \mathbb{E}Z + t\right) \leq e^{-\lambda t + v\phi(\lambda)}$. Minimizing over all $\lambda > 0$, we get $\mathbb{P}\left(Z \geq \mathbb{E}Z + t\right) \leq e^{-v\cdot h(t/v)}$ where $h(x) = (1+x)\cdot\log(1+x) - x$. $\square$

The following sub-additive increments bound can be found as Theorem 2.5 in

`Olivier Bousquet. Concentration Inequalities and Empirical Processes Theory Applied to the Analysis of Learning Algorithms. PhD thesis, Ecole Polytechnique, 2002.`

Theorem 41.3. *Let* $Z = \sup_{f \in \mathcal{F}} \sum f_i$, $\mathbb{E} f_i = 0$, $\sup_{f \in \mathcal{F}} var(f) = \sup_{f \in \mathcal{F}} \sum_{i=1}^n f_i^2 \overset{def.}{=} \sigma^2$, $\forall i \in \{1, \cdots, n\}$, $f_i \leq u \leq 1$. *Then* $\mathbb{P}\left(Z \geq \mathbb{E}Z + \sqrt{(1+u)\mathbb{E}Z + n\sigma^2 x} + \frac{x}{3}\right) \leq e^{-x}$.

Proof. Let

$$\begin{aligned}
Z &= \sup_{f \in \mathcal{F}} \sum_{i=1}^n f_i \overset{\text{def.}}{=} \sum_{i=1}^n f_i^\circ \\
Z_k &= \sup_{f \in \mathcal{F}} \sum_{i \neq k} f_i \\
Z_k' &= f_k \text{ such that } Z_k = \sup_{f \in \mathcal{F}} \sum_{i \neq k} f_i.
\end{aligned}$$

It follows that $Z_k' \leq Z - Z_k \leq u$. Let $\psi(x) = e^{-x} + x - 1$. Then

$$\begin{aligned}
e^{\lambda Z} \psi(\lambda(Z - Z_k)) &= e^{\lambda Z_k} - e^{\lambda Z} + \lambda(Z - Z_k) e^{\lambda Z} \\
&= f(\lambda)\left(Z - Z_k\right) e^{\lambda Z} + \left(\lambda - f(\lambda)\right)\left(Z - Z_k\right) e^{\lambda Z} + e^{\lambda Z_k} - e^{\lambda Z} \\
&= f(\lambda)\left(Z - Z_k\right) e^{\lambda Z} + g(Z - Z_k) e^{\lambda Z_k}.
\end{aligned}$$

In the above, $g(x) = 1 - e^{\lambda x} + (\lambda - f(\lambda)) x e^{\lambda x}$, and we define $f(\lambda) = \left(1 - e^{\lambda} + \lambda e^{\lambda}\right) / \left(e^{\lambda} + \alpha - 1\right)$ where $\alpha = 1/(1+u)$. We will need the following lemma to make use of the bound on the variance.

Lemma 41.4. *For all* $x \leq 1$, $\lambda \geq 0$ *and* $\alpha \geq \frac{1}{2}$, $g(x) \leq f(x)\left(\alpha x^2 - x\right)$.

Continuing the proof, we have

$$\begin{aligned}
e^{\lambda Z} \psi(\lambda(Z - Z_k)) &= f(\lambda)\left(Z - Z_k\right) e^{\lambda Z} + g(Z - Z_k) e^{\lambda Z_k} \\
&\leq f(\lambda)\left(Z - Z_k\right) e^{\lambda Z} + e^{\lambda Z_k} f(\lambda)\left(\alpha\left(Z - Z_k\right)^2 - \left(Z - Z_k\right)\right) \\
&\leq f(\lambda)\left(Z - Z_k\right) e^{\lambda Z} + e^{\lambda Z_k} f(\lambda)\left(\alpha\left(Z_k'\right)^2 - Z_k'\right).
\end{aligned}$$

Sum over all $k = 1, \cdots, n$, and take expectation, we get

$$\begin{aligned}
e^{\lambda Z} \sum_k \psi(\lambda(Z - Z_k)) &\leq f(\lambda) Z e^{\lambda Z} + f(\lambda) \sum_k e^{\lambda Z_k}\left(\alpha\left(Z_k'\right)^2 - Z_k'\right) \\
\mathbb{E} e^{\lambda Z} \sum_k \psi(\lambda(Z - Z_k)) &\leq f(\lambda) \mathbb{E} Z e^{\lambda Z} + f(\lambda) \sum_k \mathbb{E} e^{\lambda Z_k}\left(\alpha\left(Z_k'\right)^2 - Z_k'\right).
\end{aligned}$$

Since $\mathbb{E}Z'_k = 0$, $\mathbb{E}_{X_k}\left(Z'_k\right)^2 = \mathbb{E}f_k^2 = \mathrm{var}(f_k) \leq \sup_{f\in\mathcal{F}} \mathrm{var}(f) \leq \sigma^2$, it follows that

$$\begin{aligned}
\mathbb{E}e^{\lambda Z_k}\left(\alpha\left(Z'_k\right)^2 - Z'_k\right) &= \mathbb{E}e^{\lambda Z_k}\left(\alpha\mathbb{E}_{f_k}\left(Z'_k\right)^2 - \mathbb{E}_{f_k}Z'_k\right) \\
&\leq \alpha\sigma^2\mathbb{E}e^{\lambda Z_k} \\
&\leq \alpha\sigma^2\mathbb{E}e^{\lambda Z_k + \lambda\mathbb{E}Z'_k} \\
&\overset{\text{Jensen's inequality}}{\leq} \alpha\sigma^2\mathbb{E}e^{\lambda Z_k + \lambda Z'_k} \\
&\leq \alpha\sigma^2\mathbb{E}e^{\lambda Z}.
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{E}\left(\lambda Z e^{\lambda Z}\right) - \mathbb{E}e^{\lambda Z}\log\mathbb{E}e^{\lambda Z} &\leq \mathbb{E}e^{\lambda Z}\sum_k \psi(\lambda(Z - Z_k)) \\
&\leq f(\lambda)\mathbb{E}Ze^{\lambda Z} + f(\lambda)\alpha n\sigma^2\mathbb{E}e^{\lambda Z}.
\end{aligned}$$

Let $Z_0 = Z - \mathbb{E}$, and center Z, we get

$$\mathbb{E}\left(\lambda Z_0 e^{\lambda Z_0}\right) - \mathbb{E}e^{\lambda Z_0}\log\mathbb{E}e^{\lambda Z_0} \leq f(\lambda)\mathbb{E}Z_0e^{\lambda Z_0} + f(\lambda)\left(\alpha n\sigma^2 + \mathbb{E}Z\right)\mathbb{E}e^{\lambda Z_0}.$$

Let $F(\lambda) = \mathbb{E}e^{\lambda Z_0}$, and $\Psi(\lambda) = \log F(\lambda)$, we get

$$\begin{aligned}
(\lambda - f(\lambda))\,F'(\lambda) - F(\lambda)\log F(\lambda) &\leq f(\lambda)\left(\alpha n\sigma^2 + \mathbb{E}Z\right)F(\lambda) \\
(\lambda - f(\lambda))\underbrace{\frac{F'(\lambda)}{F(\lambda)}}_{\Psi'(\lambda)} - \underbrace{\log F(\lambda)}_{\Psi(\lambda)} &\leq f(\lambda)\left(\alpha n\sigma^2 + \mathbb{E}Z\right).
\end{aligned}$$

Solve this inequality, we get $F(\lambda) \leq e^{v\psi(-\lambda)}$ where $v = n\sigma^2 + (1+u)\mathbb{E}Z$. $\square$