Talagrand's Inequality

This document is from https://mathematicaster.org/teaching/advgraphs2025/talagrand.pdf

Throughout this document, $\Omega_1, \ldots, \Omega_n$ will be fixed probability spaces and we will consider $\Omega = \Omega_1 \times \cdots \times \Omega_n$, equipped with the product σ-algebra.

Theorem 1 (Talagrand for the working mathematician)**.** *Let $f\colon \Omega \to \mathbb{R}$ be a measurable function. Suppose that for every $\vec{x} \in \Omega$, there is another vector $\vec{\alpha}(\vec{x}) = \big(\alpha_1(\vec{x}), \ldots, \alpha_n(\vec{x})\big) \in \mathbb{R}^n$ satisfying*

$$f(\vec{x}) \leq f(\vec{y}) + \sum_{i:\ x_i \neq y_i} \alpha_i(\vec{x}) \qquad \textit{for all } \vec{y} \in \Omega.$$

For independent samples $x_1, \ldots, x_n$ with $x_i \in \Omega_i$, define $Z = f(x_1, \ldots, x_n)$. Then

$$\mathbf{Pr}\big[|Z - \mathbb{M}\, Z| \geq \lambda\big] \leq 4 \exp\left(\frac{-\lambda^2}{4 \sup_{\vec{x} \in \Omega} \|\vec{\alpha}(\vec{x})\|^2}\right).$$

Here, $\mathbb{M}\, Z$ denotes "the" median of the random variable Z.

We additionally note that, by using $-f$ instead of f, the condition can be replaced by

$$f(\vec{x}) \geq f(\vec{y}) - \sum_{i:\ x_i \neq y_i} \alpha_i(\vec{x}) \qquad \text{for all } \vec{y} \in \Omega.$$

While the above theorem is how we will generally use Talagrand's inequality, Talagrand's actually inequality is more general, and we will need to spend some time setting up the relevant defintions.

For a vector $\vec{\alpha} \in \mathbb{R}^n$, define the (signed) distance on Ω by

$$d_{\vec{\alpha}}(\vec{x}, \vec{y}) \stackrel{\text{def}}{=} \sum_{i:\ x_i \neq y_i} \alpha_i,$$

which is a weighted Hamming distance. Naturally, for a set $A \subseteq \Omega$, we extend

$$d_{\vec{\alpha}}(\vec{x}, A) \stackrel{\text{def}}{=} \inf_{\vec{a} \in A} d_{\vec{\alpha}}(\vec{x}, \vec{a}).$$

We then define *Talagrand's Convex Distance* by

$$\rho(\vec{x}, A) \stackrel{\text{def}}{=} \sup_{\|\vec{\alpha}\| = 1} d_{\vec{\alpha}}(\vec{x}, A).$$

Theorem 2 (Talagrand's Inequality)**.** *Suppose that $x_1, \ldots, x_n$ are independent samples with $x_i \in \Omega_i$ and set $\vec{x} = (x_1, \ldots, x_n)$. For any measurable $A \subseteq \Omega$,*

$$\mathbf{Pr}[\vec{x} \in A] \cdot \mathbf{Pr}[\rho(\vec{x}, A) \geq t] \leq e^{-t^2/4}.$$

We will actually prove the following:

Theorem 3. *Suppose that $x_1, \ldots, x_n$ are independent samples with $x_i \in \Omega_i$ and set $\vec{x} = (x_1, \ldots, x_n)$. For any measurable $A \subseteq \Omega$,*

$$\mathbb{E} \exp\left(\frac{\rho(\vec{x}, A)^2}{4}\right) \leq \frac{1}{\mathbf{Pr}[\vec{x} \in A]}$$

Observe that Theorem 3 implies Theorem 2. Indeed, we can use Markov's inequality with the random variable $X = \rho(\vec{x}, A)$ to bound

$$\mathbf{Pr}[X \geq t] \leq \mathbf{Pr}[X^2 \geq t^2] = \mathbf{Pr}[e^{X^2/4} \geq e^{t^2/4}] \leq \frac{\mathbb{E}\, e^{X^2/4}}{e^{t^2/4}} \leq \frac{1}{\mathbf{Pr}[\vec{x} \in A] \cdot e^{t^2/4}}.$$

In order to actually prove Theorem 3, we will need to understand Talagrand's Convex Distance in a different light. For $\vec{x} \in \Omega$ and $A \subseteq \Omega$, define

$$U(\vec{x}, A) \stackrel{\text{def}}{=} \{\vec{u} \in \{0,1\}^n : \exists \vec{a} \in A \text{ s.t. } x_i \neq a_i \implies u_i = 1\}.$$

In other words, $\vec{u} \in U(\vec{x}, A)$ if and only if there is some $\vec{y} \in A$ such that we can transform $\vec{x}$ into $\vec{y}$ by changing only coordinates for which $u_i = 1$. The set $U(\vec{x}, A)$ is a way to parameterize all "Hamming paths" from $\vec{x}$ to the set A. With this set, we can give an alternative definition of ρ by

$$\rho(\vec{x}, A) = \max_{\|\vec{\alpha}\|=1} \min_{\vec{u} \in U(\vec{x},A)} \langle \vec{\alpha}, \vec{u} \rangle.$$

Note that we can use max and min here since $U(\vec{x}, A)$ is a finite set, the sphere is compact and the map $\vec{\alpha} \mapsto \langle \vec{\alpha}, \vec{u} \rangle$ is continuous for any $\vec{u}$.

Next, define $V(\vec{x}, A) \stackrel{\text{def}}{=} \operatorname{conv}\big(U(\vec{x}, A)\big)$.

Lemma 4. *For any $\vec{x} \in \Omega$ and $A \subseteq \Omega$,*

$$\rho(\vec{x}, A) = \min_{\vec{v} \in V(\vec{x},A)} \|\vec{v}\|.$$

Note that the above minimum exists since $V(\vec{x}, A)$ is a polytope (and is hence compact) and the map $\vec{v} \mapsto \|\vec{v}\|$ is continuous.

Proof. Fix $\vec{v}^* \in V(\vec{x}, A)$ with minimum $\|\vec{v}^*\|$.

We can write $\vec{v}^* = \sum_{i=1}^k \lambda_i \vec{u}_i$ where $\lambda_1, \ldots, \lambda_k \geq 0$ and $\sum_{i=1}^k \lambda_i = 1$ and $\vec{u}_1, \ldots, \vec{u}_k \in U(\vec{x}, A)$. Fix any $\vec{\alpha}$ with $\|\vec{\alpha}\| = 1$. Then,

$$\sum_{i=1}^k \lambda_i \langle \vec{\alpha}, \vec{u}_i \rangle = \langle \vec{\alpha}, \vec{v}^* \rangle \leq \|\vec{v}^*\|,$$

where the last inequality is due to Cauchy–Schwarz. Thus, the law of averages implies that there is some $i \in [k]$ for which $\langle \vec{\alpha}, \vec{u}_i \rangle \leq \|\vec{v}^*\|$. As such, there is some $\vec{u}(\vec{\alpha}) \in U(\vec{x}, A)$ for which $\langle \vec{\alpha}, \vec{u}(\vec{\alpha}) \rangle \leq \|\vec{v}^*\|$. Hence,

$$\rho(\vec{x}, A) = \max_{\|\vec{\alpha}\|=1} \min_{\vec{u} \in U(\vec{x},A)} \leq \max_{\|\vec{\alpha}\|=1} \langle \vec{\alpha}, \vec{u}(\vec{\alpha}) \rangle \leq \|\vec{v}^*\| = \min_{\vec{v} \in V(\vec{x},A)} \|\vec{v}\|.$$

Next, we claim that $\langle \vec{a}, \vec{v}^* \rangle \geq \|\vec{v}^*\|^2$ for all $\vec{a} \in V(\vec{x}, A)$. Suppose this were note the case, so $\langle \vec{a}, \vec{v}^* \rangle \leq (1 - \epsilon)\|\vec{v}^*\|^2$ for some $\vec{a} \in A$ and some $\epsilon > 0$. For any $\lambda \in [0, 1]$, we then have

$$\begin{aligned}
\|\lambda \vec{a} + (1 - \lambda)\vec{v}^*\|^2 &= \lambda^2 \|\vec{a}\|^2 + (1 - \lambda)^2 \|\vec{v}^*\|^2 + 2\lambda(1 - \lambda)\langle \vec{a}, \vec{v}^* \rangle \\
&\leq \lambda^2 \|\vec{a}\|^2 + (1 - \lambda)^2 \|\vec{v}^*\|^2 + 2\lambda(1 - \lambda)(1 - \epsilon)\|\vec{v}^*\|^2 \\
&= \lambda^2 \|\vec{a}\|^2 + (1 - \lambda^2)\|\vec{v}^*\|^2 - 2\epsilon\lambda\|\vec{v}^*\|^2 \\
&\leq \|\vec{v}^*\| - \lambda\big(2\epsilon\|\vec{v}^*\|^2 - \lambda\|\vec{a}\|^2\big).
\end{aligned}$$

Now, since $\epsilon > 0$, we can find some $\lambda \in (0,1)$ for which $\lambda\|a\|^2 < 2\epsilon\|\vec{v}^*\|^2$. For this λ, we then have $\lambda\vec{a} + (1-\lambda)\vec{v} \in V(\vec{x}, A)$ (by convexity) and $\|\lambda\vec{a} + (1-\lambda)\vec{v}^*\|^2 < \|\vec{v}^*\|^2$; a contradiction to the definition of $\vec{v}^*$.

Now that we know that $\langle \vec{a}, \vec{v}^* \rangle \geq \|\vec{v}^*\|^2$ for all $\vec{a} \in V(\vec{x}, A)$, we can bound

$$\rho(\vec{x}, A) = \max_{\|\vec{\alpha}\|=1} \min_{\vec{u}\in U(\vec{x},A)} \langle \vec{\alpha}, \vec{u} \rangle \geq \min_{\vec{u}\in U(\vec{x},A)} \left\langle \frac{\vec{v}^*}{\|\vec{v}^*\|}, \vec{u} \right\rangle \geq \frac{\|\vec{v}^*\|^2}{\|\vec{v}^*\|} = \|\vec{v}^*\| = \min_{\vec{v}\in V(\vec{x},A)} \|\vec{v}\|. \qquad \square$$

We will also require the following two inequalities:

$$p + (1-p)e^{1/4} \leq \frac{1}{p} \qquad \text{for all } p \in (0,1], \tag{1}$$

$$\min_{\lambda\in[0,1]} e^{(1-\lambda)^2/4} r^{-\lambda} \leq 2 - r \qquad \text{for all } r \in (0,1]. \tag{2}$$

The first of these is easy to verify. Set $f(p) = p^2 + p(1-p)e^{1/4}$, so $f'(p) = e^{1/4} - 2(e^{1/4} - 1)p$. Since $e^{1/4} \geq 1$ and $e^{1/4} \leq 2 \implies 2(e^{1/4} - 1) \leq e^{1/4}$, we find that $f'(p)$ is non-negative for $p \in [0,1]$ and so f is increasing on $[0,1]$. This implies that, for $p \in [0,1]$, we have $f(p) \leq f(1) = 1$, which establishes (1).

The second is more tedious. We consider first the case when $0 < r \leq e^{-1/2}$. By selecting $\lambda = 0$, we have $\min_{\lambda\in[0,1]} e^{(1-\lambda)^2/4} r^{-\lambda} \leq e^{1/4} \leq 2 - e^{-1/2} \leq 2 - r$. We next consider the case when $e^{-1/2} < r \leq 1$. By selecting $\lambda = 1 + 2\log r$ (which is in $[0,1]$), we have $\min_{\lambda\in[0,1]} e^{(1-\lambda)^2/4} r^{-\lambda} \leq e^{\log^2 r} r^{-1-2\log r} = e^{-\log r - \log^2 r}$. Set $f(r) = e^{-\log r - \log^2 r}$, so $f(1) = 1$, $f'(1) = -1$ and $f''(x) = 2x^{-3-\log x}(2\log x + 3)\log x$. Now, observe that, for $x \in (0,1]$, $f''(x) > 0 \implies 2\log x < -3 \implies x < e^{-3/2} < e^{-1/2}$. Thus, $f''(x) \leq 0$ for all $x \in [r,1]$ since $r > e^{-1/2}$. Now, Taylor's theorem tells us that there is some $x \in [r,1]$ for which

$$f(r) = f(1) + f'(1)(r-1) + \frac{f''(x)}{2}(r-1)^2 = 2 - r + \frac{f''(x)}{2}(r-1)^2.$$

Thus, since $f''(x) \leq 0$ for all $x \in [r,1]$, we have $f(r) \leq 2 - r$ as needed to establish (2).

Additionally, we will need the following famous inequality:

Theorem 5 (Hölder's Inequality)**.** *Let X, Y be non-negative random variables. For any $\lambda \in [0,1]$,*

$$\mathbb{E}\big[X^\lambda Y^{1-\lambda}\big] \leq \big(\mathbb{E}\, X\big)^\lambda \big(\mathbb{E}\, Y\big)^{1-\lambda}.$$

By setting $\lambda = 1/2$, we recover the Cauchy–Schwarz inequality.

Proof. We begin by noting that for any $a, b > 0$,

$$\log\big(a^\lambda b^{1-\lambda}\big) = \lambda \log a + (1-\lambda)\log b \leq \log\big(\lambda a + (1-\lambda)b\big),$$

where the inequality follows from the concavity of log. Therefore,

$$a^\lambda b^{1-\lambda} \leq \lambda a + (1-\lambda)b.$$

Note that this inequality trivially holds if either a or b is 0, and so the inequality holds for all $a, b \geq 0$. Applying this with $a = X/\mathbb{E}\, X$ and $b = Y/\mathbb{E}\, Y$ then yields

$$\left(\frac{X}{\mathbb{E}\, X}\right)^\lambda \left(\frac{Y}{\mathbb{E}\, Y}\right)^{1-\lambda} \leq \lambda \frac{X}{\mathbb{E}\, X} + (1-\lambda)\frac{Y}{\mathbb{E}\, Y}$$

$$\implies \mathbb{E}\left[\left(\frac{X}{\mathbb{E}\, X}\right)^\lambda \left(\frac{Y}{\mathbb{E}\, Y}\right)^{1-\lambda}\right] \leq \lambda + (1-\lambda) = 1.$$

Multiplying both sides by $\big(\mathbb{E}\, X\big)^\lambda \big(\mathbb{E}\, Y\big)^{1-\lambda}$ then yields the claim. $\square$

Finally, we will rely on Fubini's theorem throughout the proof, though we will not give a proof here.

Proof of Theorem 3. We prove the claim by induction on n.

When $n = 1$, observe that $\rho(x, A) = 0$ if $x \in A$ and $\rho(x, A) = 1$ if $x \notin A$. Therefore, we can use (1) to bound

$$\mathbb{E} \exp\left(\frac{\rho(x, A)^2}{4}\right) = \mathbf{Pr}[x \in A] + (1 - \mathbf{Pr}[x \in A])e^{1/4} \leq \frac{1}{\mathbf{Pr}[x \in A]}.$$

We now suppose that $n \geq 2$, so fix a measurable $A \subseteq \Omega_1 \times \cdots \times \Omega_n$. For $y \in \Omega_n$, define $A_y \stackrel{\text{def}}{=} \{\vec{x} \in \Omega_1 \times \cdots \times \Omega_{n-1} : (\vec{x}, y) \in A\}$ and $B = \bigcup_{y \in \Omega_n} A_y$. Fubini's theorem implies that B is measurable.

We will prove that for any fixed $y \in \Omega_n$

$$\mathop{\mathbb{E}}_{\vec{x}} \exp\left(\frac{\rho\big((\vec{x}, y), A\big)^2}{4}\right) \leq \frac{1}{\mathbf{Pr}[\vec{x} \in B]}\left(2 - \frac{\mathbf{Pr}[\vec{x} \in A_y]}{\mathbf{Pr}[\vec{x} \in B]}\right) \tag{3}$$

Indeed, if this holds, then we can apply Fubini's theorem to bound

$$\begin{aligned}
\mathop{\mathbb{E}}_{\vec{x},y} \exp\left(\frac{\rho\big((\vec{x}, y), A\big)^2}{4}\right) &\leq \mathop{\mathbb{E}}_{y} \frac{1}{\mathbf{Pr}[\vec{x} \in B]}\left(2 - \frac{\mathbf{Pr}[\vec{x} \in A_y]}{\mathbf{Pr}[\vec{x} \in B]}\right) = \frac{1}{\mathbf{Pr}[\vec{x} \in B]}\left(2 - \frac{\mathbb{E}_y \mathbf{Pr}[\vec{x} \in A_y]}{\mathbf{Pr}[\vec{x} \in B]}\right) \\
&= \frac{1}{\mathbf{Pr}[\vec{x} \in B]}\left(2 - \frac{\mathbf{Pr}[(\vec{x}, y) \in A]}{\mathbf{Pr}[\vec{x} \in B]}\right) \\
&= \frac{1}{\mathbf{Pr}[(\vec{x}, y) \in A]} \cdot \frac{\mathbf{Pr}[(\vec{x}, y) \in A]}{\mathbf{Pr}[\vec{x} \in B]}\left(2 - \frac{\mathbf{Pr}[(\vec{x}, y) \in A]}{\mathbf{Pr}[x \in B]}\right) \\
&\leq \frac{1}{\mathbf{Pr}[(\vec{x}, y) \in A]},
\end{aligned}$$

where the last inequality follows from the AM–GM inequality since $(\vec{x}, y) \in A \implies \vec{x} \in B$.

Thus, we turn our attention to proving (3). Consider $(\vec{x}, y) \in \Omega$. In order to transform $(\vec{x}, y)$ into an element of A, we could either change $\vec{x}$ to an element of B and then change y appropriately, or we could leave y fixed and change $\vec{x}$ to an element of A_y. This implies that

$$\begin{aligned}
\vec{s} \in U(\vec{x}, B) &\implies (\vec{s}, 1) \in U\big((\vec{x}, y), A\big), \\
\vec{t} \in U(\vec{x}, A_y) &\implies (\vec{s}, 0) \in U\big((\vec{x}, y), A\big).
\end{aligned}$$

By then taking convex combinations, we find that for any $\vec{s} \in V(\vec{x}, B)$ and $\vec{t} \in V(\vec{x}, A_y)$ and $\lambda \in [0, 1]$, we have

$$\big((1 - \lambda)\vec{s} + \lambda\vec{t}, 1 - \lambda\big) \in V\big((\vec{x}, y), A\big).$$

Due to Lemma 4, we can select $\|\vec{s}\| = \rho(\vec{x}, B)$ and $\|\vec{t}\| = \rho(\vec{x}, A_y)$ and have

$$\begin{aligned}
\rho\big((\vec{x}, y), A\big)^2 &\leq \big\|\big((1 - \lambda)\vec{s} + \lambda\vec{t}, 1 - \lambda\big)\big\|^2 = \|(1 - \lambda)\vec{s} + \lambda\vec{t}\|^2 + (1 - \lambda)^2 \\
&\leq (1 - \lambda)\|\vec{s}\|^2 + \lambda\|\vec{t}\|^2 + (1 - \lambda)^2 \\
&= (1 - \lambda)\rho(\vec{x}, B)^2 + \lambda\rho(\vec{x}, A_y)^2 + (1 - \lambda)^2,
\end{aligned}$$

where the inequality in the second line follows from the fact that $\vec{v} \mapsto \|\vec{v}\|^2$ is a convex function.

We now combine the above inequality with Hölder's inequality (Theorem 5) and the induction hypothesis to bound

$$\begin{aligned}
\mathop{\mathbb{E}}_{\vec{x}} \exp\left(\frac{\rho\big((\vec{x},y),A\big)^2}{4}\right) &\le \mathop{\mathbb{E}}_{\vec{x}} \exp\left(\frac{(1-\lambda)\rho(\vec{x},B)^2 + \lambda\rho(\vec{x},A_y)^2 + (1-\lambda)^2}{4}\right) \\
&= e^{(1-\lambda)^2/4} \mathop{\mathbb{E}}_{\vec{x}}\left[\left(e^{\rho(\vec{x},B)^2/4}\right)^{1-\lambda}\left(e^{\rho(\vec{x},A_y)^2/4}\right)^{\lambda}\right] \\
&\le e^{(1-\lambda)^2/4}\left(\mathop{\mathbb{E}}_{\vec{x}} e^{\rho(\vec{x},B)^2/4}\right)^{1-\lambda}\left(\mathop{\mathbb{E}}_{\vec{x}} e^{\rho(\vec{x},A_y)^2/4}\right)^{\lambda} \\
&\le e^{(1-\lambda)^2/4}\left(\frac{1}{\mathbf{Pr}[\vec{x}\in B]}\right)^{1-\lambda}\left(\frac{1}{\mathbf{Pr}[\vec{x}\in A_y]}\right)^{\lambda} \\
&= \frac{1}{\mathbf{Pr}[\vec{x}\in B]}\cdot e^{(1-\lambda)^2/4}\left(\frac{\mathbf{Pr}[\vec{x}\in A_y]}{\mathbf{Pr}[\vec{x}\in B]}\right)^{-\lambda}.
\end{aligned}$$

Finally, $A_y \subseteq B$ and so we can apply (2) to finally establish (3) as needed. □

Now that we have proved Talagrand's inequality, we will will establish the version mentioned at the beginning of this document.

Proof of Theorem 1. Set $K = \sup_{\vec{x}\in\Omega}\|\vec{\alpha}(\vec{x})\|$ for the proof.

Fix $r \in \mathbb{R}$ to be chosen later and define $A \stackrel{\text{def}}{=} \{Z \le r - \lambda\}$.

Now, fix any $\vec{x} \in \Omega$ with $f(\vec{x}) \ge r$. By assumption,

$$d_{\vec{\alpha}(\vec{x})}(\vec{x},A) = \inf_{\vec{y}\in A} d_{\vec{\alpha}(\vec{x})}(\vec{x},\vec{y}) = \inf_{\vec{y}\in A}\sum_{i:\ x_i\neq y_i}\alpha_i(\vec{x}) \ge \inf_{\vec{y}\in A}\big(f(\vec{x}) - f(\vec{y})\big) \ge r - (r-\lambda) = \lambda.$$

Therefore,

$$\rho(\vec{x},A) \ge \frac{d_{\vec{\alpha}(\vec{x})}(\vec{x},A)}{\|\vec{\alpha}(\vec{x})\|} \ge \frac{\lambda}{K} \qquad \text{whenever } f(\vec{x}) \ge r.$$

Using this fact, we can finally apply Talagrand's inequality (Theorem 2) to bound

$$\mathbf{Pr}[Z \le r-\lambda]\cdot\mathbf{Pr}[Z \ge r] = \mathbf{Pr}[\vec{x}\in A]\cdot\mathbf{Pr}[f(\vec{x}) \ge r] \le \mathbf{Pr}[\vec{x}\in A]\cdot\mathbf{Pr}[\rho(\vec{x},A) \ge \lambda/K] \le e^{-\lambda^2/4K^2}.$$

Selecting $r = \mathbb{M}\,Z$ then yields,

$$e^{-\lambda^2/4K^2} \ge \mathbf{Pr}[Z \le \mathbb{M}\,Z - \lambda]\cdot\mathbf{Pr}[Z \ge \mathbb{M}\,Z] \ge \frac{1}{2}\,\mathbf{Pr}[Z \le \mathbb{M}\,Z - \lambda].$$

Similarly, selecting $r = \mathbb{M}\,Z + \lambda$ then yields,

$$e^{-\lambda^2/4K^2} \ge \mathbf{Pr}[Z \le \mathbb{M}\,Z]\cdot\mathbf{Pr}[Z \ge \mathbb{M}\,Z + \lambda] \ge \frac{1}{2}\,\mathbf{Pr}[Z \ge \mathbb{M}\,Z + \lambda].$$

Thus, the union bound tells us that

$$\mathbf{Pr}\big[|Z - \mathbb{M}\,Z| \ge \lambda\big] \le \mathbf{Pr}[Z \le \mathbb{M}\,Z - \lambda] + \mathbf{Pr}[Z \ge \mathbb{M}\,Z + \lambda] \le 4e^{-\lambda^2/4K^2}.$$ □

Talagrand's inequality implies a weak form of McDiarmid's inequality, which, for most of our applications, would be perfectly sufficient.

Theorem 6 (Weak McDiarmid)**.** *Let* $f\colon \Omega \to \mathbb{R}$ *be a measurable function such that* $|f(\vec{x}) - f(\vec{y})| \le c_i$ *when* $\vec{x}$ *and* $\vec{y}$ *differ only in the* i*th coordinate. Set* $K = \sqrt{\sum_{i=1}^n c_i^2}$*. For independent samples* $x_1, \ldots, x_n$ *with* $x_i \in \Omega_i$*, define* $Z = f(x_1, \ldots, x_n)$*. Then,*

$$\mathbf{Pr}\big[|Z - \mathbb{E}\, Z| \ge \lambda\big] \le 4e^{-\lambda^2/16K^2} \qquad \text{for every } \lambda > 8K.$$

Recall that McDiarmid's inequality states that actually

$$\mathbf{Pr}\big[|Z - \mathbb{E}\, Z| \ge \lambda\big] \le 2e^{-2\lambda^2/K^2} \qquad \text{for every } \lambda > 0.$$

Usually, the constant factors do not matter much and we consider $\lambda \gg K$, so the weak form of McDiarmid usually suffices for our purposes.

In order to prove Theorem 6, we begin with a fairly simple observation: If a random variable has subGaussian tails centered at m, then $\mathbb{E}\, X$ and m are not too far apart.

Lemma 7. *Let* X *be a random variable and fix* $m \in \mathbb{R}$*. Suppose that there exist* $C, c > 0$ *such that*

$$\mathbf{Pr}\big[|X - m| \ge \lambda\big] \le Ce^{-c\lambda^2} \qquad \text{for every } \lambda > 0.$$

Then $|\mathbb{E}\, X - m| \le \sqrt{C/c}$*.*

Proof. Using Jensen's inequality, we compute,

$$\begin{aligned}\big(\mathbb{E}\, X - m\big)^2 = \big(\mathbb{E}(X - m)\big)^2 \le \mathbb{E}(X-m)^2 &= \int_0^\infty \mathbf{Pr}[(X-m)^2 \ge t]dt \\ &= \int_0^\infty \mathbf{Pr}\big[|X - m| \ge \sqrt{t}\big]dt \le \int_0^\infty Ce^{-ct}dt = \frac{C}{c}\int_0^\infty e^{-t}dt = \frac{C}{c}.\end{aligned}$$

□

We can now prove that Talagrand implies the weak form of McDiarmid.

Proof of Theorem 6. For each $\vec{x} \in \Omega$, define $\vec{\alpha}(\vec{x}) = (c_1, \ldots, c_n)$. Then, by the assumption and the triangle inequality,

$$|f(\vec{y}) - f(\vec{x})| \le \sum_{i\colon\, y_i \ne x_i} c_i \implies f(\vec{x}) \le f(\vec{y}) + \sum_{i\colon\, x_i \ne y_i} \alpha_i(\vec{x}).$$

Furthermore, $\sup_{\vec{x}\in\Omega}\|\vec{\alpha}(\vec{x})\| = K$. Thus, Talagrand's inequality (Theorem 1) implies that

$$\mathbf{Pr}\big[|Z - \mathbb{M}\, Z| \ge \lambda\big] \le 4e^{-\lambda^2/4K^2} \qquad \text{for every } \lambda > 0.$$

Thus, we can apply Lemma 7 to find that $|\mathbb{M}\, Z - \mathbb{E}\, Z| \le \sqrt{\frac{4}{1/4K^2}} = 4K$. Now, note that $\lambda - 4K \ge \lambda/2$ since $\lambda \ge 8K$, so we can additionally use the triangle inequality to bound

$$\begin{aligned}\mathbf{Pr}\big[|Z - \mathbb{E}\, Z| \ge \lambda\big] &\le \mathbf{Pr}\big[|Z - \mathbb{M}\, Z| + |\mathbb{M}\, Z - \mathbb{E}\, Z| \ge \lambda\big] \\ &\le \mathbf{Pr}\big[|Z - \mathbb{M}\, Z| > \lambda - 4K\big] \le 4e^{-(\lambda-4K)^2/4K^2} \\ &\le 4e^{-(\lambda/2)^2/4K^2} = 4e^{-\lambda^2/16K^2}.\end{aligned}$$

□