Talagrand’s Inequality

Fall 2019 Reading Course: Concentration of Measure
University of Utah

December 3, 2019
References


Notation + Setup

Throughout: $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ has iid components.\(^1\)

Norms and distances: Euclidean in $\mathbb{R}^n$

We let the support of $X$ be $S \subset \mathbb{R}^n$, which is a product set $S = \times_{j=1}^n S_1$.

$A \subset \mathbb{R}^n$ is $X$-measurable.

Simplified example: $X_j$ is a Rademacher RV, $S = \{-1, 1\}^n$, $A$ is a subspace of $\mathbb{R}^n$.

\(^1\)Identical distribution is not necessary, but we will assume it for simplicity.
Throughout: \( X = (X_1, \ldots, X_n) \in \mathbb{R}^n \) has iid components.\(^1\)

Norms and distances: Euclidean in \( \mathbb{R}^n \)

We let the support of \( X \) be \( S \subset \mathbb{R}^n \), which is a product set \( S = \times_{j=1}^n S_1 \).

\( A \subset \mathbb{R}^n \) is \( X \)-measurable.

Simplified example: \( X_j \) is a Rademacher RV, \( S = \{-1, 1\}^n \), \( A \) is a subspace of \( \mathbb{R}^n \).

Overall goal: to study the probability that \( X \) is “far” from \( A \).

\(^1\)Identical distribution is not necessary, but we will assume it for simplicity.
Talagrand’s Inequality (1/2)

We consider a simplified setup: assume that $A$ is convex and $X$ has iid Rademacher components.

Then for all $t > 0$:

$$P [X \in A] P [\text{dist}(X, A) \geq t] \leq \exp(-ct^2).$$

(babyTI)

For example, $c = 1/16$ works.

Virtues:

- dimension-independent constants
- probability that $X$ is far from $A$ is exponentially small, if $A$ is “large” enough
Talagrand’s Inequality (2/2)

The general setup is more involved: “distance” needs to be properly defined to work out.

Let $A$ be a general set, and let $\alpha \in \mathbb{R}_{+}^{n}$.

Consider the $\alpha$-weighted Hamming distance between $x$ and $A$:

$$d_{\alpha}(x, A) = \inf_{y \in A} \| [x - y] \|_{1, \alpha},$$

where

$$[x - y] := (\mathbb{1}_{x_1 \neq y_1}, \mathbb{1}_{x_2 \neq y_2}, \ldots, \mathbb{1}_{x_n \neq y_n}) \in \{0, 1\}^{n}, \quad \| x \|_{1, \alpha} := \sum_{j=1}^{n} \alpha_j |x_j|$$

Talagrand’s convex distance is defined as:

$$\rho(x, A) := \sup_{\| \alpha \| = 1} d_{\alpha}(x, A),$$

Talagrand’s inequality states:

$$P [X \in A] P [\rho(X, A) \geq t] \leq \exp(-\frac{1}{4}t^2). \quad \forall \ t > 0 \quad (TI)$$
Applications (1a/2)

Concentration of Lipschitz functions: let $f : S \to \mathbb{R}$ be 1-Lipschitz:

$$|f(x) - f(y)| \leq \|x - y\|_1.$$

Assume $S$ is bounded, so that $d := \max_{a, b \in S_1} |a - b| < \infty$. Then $\tilde{f} := f/(d\sqrt{n})$ satisfies:

$$\left| \tilde{f}(x) - \tilde{f}(y) \right| \leq \frac{\|x - y\|_1}{d\sqrt{n}} \leq \|x - y\|_{1, \tilde{\alpha}} = d\tilde{\alpha}(x, y)$$

where $\tilde{\alpha} = \frac{1}{\sqrt{n}} (1, 1, \ldots, 1) \in \mathbb{R}^n$ has Euclidean norm 1.

Given some $m \in \mathbb{R}$, let $A := \tilde{f}^{-1}((-\infty, m])$. Talagrand’s inequality states:

$$P \left[ \tilde{f}(X) \leq m \right] P \left[ \rho(X, A) \geq t \right] \leq \exp(-t^2/4)$$
Applications (1b/2)

\[ P \left[ \tilde{f}(X) \leq m \right] P \left[ \rho(X, A) \geq t \right] \leq \exp(-t^2/4) \]  

(1)

Suppose now that \( x \) is such that \( \tilde{f}(x) \geq m + t \). Then for any \( y \in A \):

\[ m + t \leq \tilde{f}(x) \leq \tilde{f}(y) + d_\alpha(x, y) \quad \implies \quad m + t \leq m + d_\alpha(x, A) \leq m + \rho(x, A) \]

Thus: \( \tilde{f}^{-1}([m + t, \infty)) \subseteq \{x \in S \mid \rho(x, A) \geq t\} \).

Then (1) implies:

\[ P \left[ \tilde{f}(X) \leq m \right] P \left[ \tilde{f}(X) \geq m + t \right] \leq \exp(-t^2/4) \]

Take \( m = \text{med}(f(X)) \) and \( m = t - \text{med}(f(X)) \) to yield:

\[ P \left[ |\tilde{f}(X) - \text{med}(\tilde{f}(X))| \geq t \right] \leq 4 \exp(-t^2/4), \]
Applications (2/2)

Large deviation inequality: Let $V$ be a subspace of $\mathbb{R}^n$. For $r > 0$, define

$$A = A(r) = \{x \mid \text{dist}(x, V) \leq r\}$$

Consider $X$ as an $n$-sequence of iid Rademacher RV's. Then for any $t > 0$, (babyTI) implies

$$P[\text{dist}(X, V) \leq r] P[\text{dist}(X, V) \geq r + t] \leq \exp(-ct^2).$$

Let $M := \text{med}(\text{dist}(X, V))$.

Picking $r = M$, $r = M - t$:

$$P[|\text{dist}(X, V) - M| \geq t] \leq 4 \exp(-ct^2).$$

Here, if $d = \dim V$, then $M \sim \sqrt{n - d}$. 

Talagrand’s Inequality
BabyTI proof (1/4)

A convex, $X \in \{-1, 1\}^n$ iid Rademacher sequence.

Ideas:
- inequalities: Jensen, Markov, Hölder
- induction on dimension $n$
- convexity of $A$: conditioning for inductive step

Markov’s inequality,

$$P(|X| \geq t) \leq \frac{1}{\phi(t)} \mathbb{E}\phi(|X|), \quad \phi : \mathbb{R}_+ \to \mathbb{R} \text{ monotonically increasing}$$

applied to $\phi(r) = \exp(cr)$ reveals that

$$P(X \in A)\mathbb{E}\left(\exp(c\text{dist}^2(X, A))\right) \leq 1 \implies \text{(babyTI)},$$

so we focus on proving this.
BabyTI proof (2/4)

\[ P(X \in A) \mathbb{E} \left( \exp(\text{cdist}^2(X, A)) \right) \leq 1 \implies \text{(babyTI)}, \]

Induction: \( n = 0 \) is easy. Now assume this is true for \( n - 1 \). \( X \in \{-1, 1\}^n \).

We compute two components: (i) \( P(X \in A) \) and (ii) \( \mathbb{E} \left( \exp(\text{cdist}^2(X, A)) \right) \)

(i): Decompose \( X = (\tilde{X}, x_n) \), for \( x_n = \pm 1 \). Fix \( t \in \mathbb{R} \):

\[ A_t := \{ \tilde{x} \in \mathbb{R}^{n-1} \mid (\tilde{x}, t) \in A \}. \]

Note: \( A_t \) is convex since \( A \) is convex.

Then:

\[ p := P(X \in A) = \frac{1}{2} P(\tilde{X} \in A_{-1}) + \frac{1}{2} P(\tilde{X} \in A_{+1}). \]

Symmetrize around \( p \):

\[ p_\pm := P(\tilde{X} \in A_{\pm 1}) =: p(1 \pm q), \quad q \in [0, 1] \]
(ii): Define $\tilde{Y}_\pm = \tilde{Y}(\tilde{X})$ as the closest point in $A_\pm$ to $\tilde{X}$, i.e.,

$$\|\tilde{X} - Y_\pm\| = \text{dist}(\tilde{X}, A_\pm).$$

(2)

Then:

$$(\tilde{Y}_\pm, \pm 1) \in A \quad \text{convex} \quad \iff \quad (1 - \lambda)(\tilde{Y}_{X_n}, X_n) + \lambda(\tilde{Y}_- X_n, -X_n) \in A.$$

Thus:

$$\text{dist}^2(X, A) \leq \left\| (1 - \lambda)(\tilde{Y}_{X_n}, X_n) + \lambda(\tilde{Y}_- X_n, -X_n) - X \right\|^2$$

(Pythagoras)

$$\leq 4\lambda^2 + \left\| (1 - \lambda)(\tilde{Y}_{X_n} - \tilde{X}) + \lambda(\tilde{Y}_- X_n - \tilde{X}) \right\|^2$$

(Jensen)

$$\leq 4\lambda^2 + (1 - \lambda) \left\| \tilde{Y}_{X_n} - \tilde{X} \right\|^2 + \lambda \left\| \tilde{Y}_- X_n - \tilde{X} \right\|^2$$

(2)

$$\leq 4\lambda^2 + (1 - \lambda)\text{dist}^2(\tilde{X}, A_{X_n}) + \lambda \text{dist}^2(\tilde{X}, A_{-X_n})$$

Talagrand's Inequality
BabyTI proof (4/4)

Exponentiating and $\tilde{X}$-expectation’ing,

$$
\mathbb{E}_{\tilde{X}} \exp(c \text{dist}^2(X, A)) \leq e^{4c\lambda^2} \mathbb{E}_{\tilde{X}} \left[ \left( e^{c \text{dist}^2(\tilde{X}, A_{X_n})} \right)^{1-\lambda} \left( e^{c \text{dist}^2(\tilde{X}, A_{-X_n})} \right)^{\lambda} \right]
$$

Recall: $P(\tilde{X} \in A_{X_n}) = p_{+1}$, and inductive hypothesis:

$$
P(\tilde{X} \in A_{X_n}) \mathbb{E}_{\tilde{X}} \exp(c \text{dist}^2(\tilde{X}, A_{X_n})) \leq 1,
$$

Therefore:

$$
\mathbb{E}_{\tilde{X}} \exp(c \text{dist}^2(X, A)) \leq e^{4c\lambda^2} \mathbb{E}_{\tilde{X}} \left[ \left( e^{c \text{dist}^2(\tilde{X}, A_{X_n})} \right)^{1-\lambda} \left( e^{c \text{dist}^2(\tilde{X}, A_{-X_n})} \right)^{\lambda} \right]
$$

[Hölder]

$$
\leq e^{4c\lambda^2} \left[ \mathbb{E}_{\tilde{X}} \left( e^{c \text{dist}^2(\tilde{X}, A_{X_n})} \right) \right]^{1-\lambda} \left[ \mathbb{E}_{\tilde{X}} e^{c \text{dist}^2(\tilde{X}, A_{-X_n})} \right]^{\lambda}
$$

$$
\leq e^{4c\lambda^2} \left( \frac{1}{p_{X_n}} \right)^{1-\lambda} \left( \frac{1}{p_{-X_n}} \right)^{\lambda}.
$$

Then if $X_n = \pm 1$, we have

$$
p \mathbb{E}_{\tilde{X}} \exp(c \text{dist}^2(X, A)) \leq e^{4c\lambda^2} \frac{1}{(1 \pm q)^{1-\lambda}(1 \mp q)^{\lambda}} \leq 1
$$

Final step: expectation wrt $X_n$, taking $c$ small enough ensures that $\lambda$ can be chosen to make the inequality work.