Chapter 2.15: Concentration

Amine Hadji

University of Leiden

June 22, 2020

Last week...

... Bart answered two "important" questions

Last week...

... Bart answered two "important" questions

- how to bound $E^*\|\mathbb{G}_n\|_{\mathcal{F}}$
- how to bound $P^*(\|\mathbb{G}_n\|_{\mathcal{F}} \geq t)$

Last week...

... Bart answered two "important" questions

- how to bound $E^*\|\mathbb{G}_n\|_{\mathcal{F}}$
- how to bound $P^*(\|\mathbb{G}_n\|_{\mathcal{F}} \geq t)$

This week, we are focusing on finding a bound for $P^*(|\|\mathbb{G}_n\|_{\mathcal{F}} - E^*\|\mathbb{G}_n\|_{\mathcal{F}}| \geq t)$.

Overview

1. McDiarmid's inequality
2. Properties of entropies
3. Talagrand's Inequality

Bounded differences functions

Definition

Consider T a measurable function of independent variables $X_1, ..., X_n$. We say that T satisfies the *bounded difference inequality* if there exist $c_1, ..., c_n > 0$ such that:

$$|T(x_1, ..., x_n) - T(x_1, ..., x_{i-1}, x_i', x_{i+1}, ..., x_n)| \leq c_i$$

for every $x_1, ..., x_n, x_i' \in \mathcal{X}$, for all $i = 1, ..., n$.

McDiarmid's inequality

Proposition (McDiarmid's inequality)

If T is a measurable function of independent variables $X_1, \ldots, X_n$ which satisfies the bounded difference inequality almost surely, then for all $t > 0$

$$P(T - ET \geq t) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^n c_i^2}\right),$$

$$P(T - ET \leq -t) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^n c_i^2}\right).$$

Reminder

Theorem (Hoeffding's inequality)

Let X a bounded random variable $X \in [a, b]$ a.s. such that $EX = 0$, then for all $\lambda > 0$

$$Ee^{\lambda X} \leq \exp(\lambda^2 (b - a)^2 / 8).$$

McDiarmid's inequality

Proof:

- Let $T_i := E(T|X_1, ..., X_i) - E(T|X_1, ..., X_{i-1})$

McDiarmid's inequality

Proof:

- Let $T_i := E(T|X_1, ..., X_i) - E(T|X_1, ..., X_{i-1})$
- $T - ET = \sum_{i=1}^{n} T_i$

McDiarmid's inequality

Proof:

- Let $T_i := E(T|X_1, ..., X_i) - E(T|X_1, ..., X_{i-1})$
- $T - ET = \sum_{i=1}^{n} T_i$
- Let $\lambda > 0$, such that
  $P(T - ET \geq t) = P(\exp(\lambda \sum_{i=1}^{n} T_i) \geq e^{\lambda t})$

McDiarmid's inequality

Proof:

- Let $T_i := E(T|X_1, ..., X_i) - E(T|X_1, ..., X_{i-1})$
- $T - ET = \sum_{i=1}^{n} T_i$
- Let $\lambda > 0$, such that $P(T - ET \geq t) = P(\exp(\lambda \sum_{i=1}^{n} T_i) \geq e^{\lambda t})$
- Using Markov's inequality $P(T - ET \geq t) \leq e^{-\lambda t} E \exp(\lambda \sum_{i=1}^{n} T_i)$

McDiarmid's inequality

- Using Markov's inequality
  $P(T - ET \geq t) \leq e^{-\lambda t} E \exp(\lambda \sum_{i=1}^{n} T_i)$

Now, we would like to use Hoeffding's inequality.

- $E(T_i | X_1, ..., X_{i-1}) = 0$
- $T_i \in [A_i, B_i]$
  - $A_i = \inf_x E(T | X_1, ..., X_{i-1}, x) - E(T | X_1, ..., X_{i-1})$
  - $B_i = \sup_x E(T | X_1, ..., X_{i-1}, x) - E(T | X_1, ..., X_{i-1})$
  - $B_i - A_i \leq c_i$

McDiarmid's inequality

- Using Markov's inequality
  $P(T - ET \geq t) \leq e^{-\lambda t} E\exp(\lambda \sum_{i=1}^{n} T_i)$

Now, we would like to use Hoeffding's inequality.

- $E(T_i|X_1, ..., X_{i-1}) = 0$
- $T_i \in [A_i, B_i]$
  - $A_i = \inf_x E(T|X_1, ..., X_{i-1}, x) - E(T|X_1, ..., X_{i-1})$
  - $B_i = \sup_x E(T|X_1, ..., X_{i-1}, x) - E(T|X_1, ..., X_{i-1})$
  - $B_i - A_i \leq c_i$
- $E(\exp(\lambda T_i)|X1, ..., X_{i-1}) \leq e^{\lambda^2 c_i^2/8}$
- Applying this inequality n times gives
  $P(T - ET \geq t) \leq e^{-\lambda t}\exp(\lambda^2 \sum_{i=1}^{n} c_i^2/8)$
- $P(T - ET \geq t) \leq \inf_{\lambda>0} e^{-\lambda t}\exp(\lambda^2 \sum_{i=1}^{n} c_i^2/8) = \exp\left(\frac{-2t^2}{\sum_{i=1}^{n} c_i^2}\right)$

Bounded difference theorems

Do $\|\mathbb{G}_n\|_{\mathcal{F}}$ and $\sup_{f\in\mathcal{F}} \mathbb{G}_n f$ satisfy the bounded difference inequality?

Bounded difference theorems

Do $\|\mathbb{G}_n\|_{\mathcal{F}}$ and $\sup_{f\in\mathcal{F}} \mathbb{G}_n f$ satisfy the bounded difference inequality?
In general, no.

Bounded difference theorems

Do $\|\mathbb{G}_n\|_{\mathcal{F}}$ and $\sup_{f\in\mathcal{F}}\mathbb{G}_n f$ satisfy the bounded difference inequality?
In general, no. However,

Theorem

If $\mathcal{F}$ is a class of measurable functions such that for every $f:\mathcal{X}\to\mathbb{R}\in\mathcal{F}$ and every $x,y\in\mathcal{X}$, $|f(x)-f(y)|\leq 1$, then for all $t\geq 0$

$$P^*\left(|\|\mathbb{G}_n\|_{\mathcal{F}} - E^*\|\mathbb{G}_n\|_{\mathcal{F}}| \geq t\right) \leq 2e^{-2t^2}$$

$$P^*\left(|\sup_{f\in\mathcal{F}}\mathbb{G}_n f - E^*\sup_{f\in\mathcal{F}}\mathbb{G}_n f| \geq t\right) \leq 2e^{-2t^2}$$

The entropy

The *entropy* of a $Y > 0$ a.s. is $Ent(Y) = EY \log Y - EY \log(EY)$. It satisfies:

The entropy

The *entropy* of a $Y > 0$ a.s. is $Ent(Y) = EY \log Y - EY \log(EY)$. It satisfies:

- (i) $Ent(Y) \geq 0$
- (ii) $Ent(\gamma Y) = \gamma Ent(Y)$ for any $\gamma \in \mathbb{R}$

The entropy

The *entropy* of a $Y > 0$ a.s. is $Ent(Y) = EY\log Y - EY\log(EY)$. It satisfies:

- (i) $Ent(Y) \geq 0$
- (ii) $Ent(\gamma Y) = \gamma Ent(Y)$ for any $\gamma \in \mathbb{R}$
- (iii) $Ent(Y) = \sup\{EYU : Ee^U = 1\}$
- (iv) $Ent(Y) = \sup\{EY\log(V/EV) : V > 0\}$

The entropy

The *entropy* of a $Y > 0$ a.s. is
$Ent(Y) = EY \log Y - EY \log(EY)$. It satisfies:

- (i) $Ent(Y) \geq 0$
- (ii) $Ent(\gamma Y) = \gamma Ent(Y)$ for any $\gamma \in \mathbb{R}$
- (iii) $Ent(Y) = \sup\{EYU : Ee^U = 1\}$
- (iv) $Ent(Y) = \sup\{EY \log(V/EV) : V > 0\}$
- (v) $Ent(Y) = \inf_{r>0} E(Y \log Y - (1 + \log r)Y + r)$

The entropy

The *entropy* of a $Y > 0$ a.s. is $Ent(Y) = EY \log Y - EY \log(EY)$. It satisfies:

- (i) $Ent(Y) \geq 0$
- (ii) $Ent(\gamma Y) = \gamma Ent(Y)$ for any $\gamma \in \mathbb{R}$
- (iii) $Ent(Y) = \sup\{EYU : Ee^U = 1\}$
- (iv) $Ent(Y) = \sup\{EY \log(V/EV) : V > 0\}$
- (v) $Ent(Y) = \inf_{r>0} E(Y \log Y - (1 + \log r)Y + r)$
- (vi) The function $Ent(.)$ is convex

The entropy

The *entropy* of a $Y > 0$ a.s. is
$Ent(Y) = EY \log Y - EY \log(EY)$. It satisfies:

- (i) $Ent(Y) \geq 0$
- (ii) $Ent(\gamma Y) = \gamma Ent(Y)$ for any $\gamma \in \mathbb{R}$
- (iii) $Ent(Y) = \sup\{EYU : Ee^{U} = 1\}$
- (iv) $Ent(Y) = \sup\{EY \log(V/EV) : V > 0\}$
- (v) $Ent(Y) = \inf_{r>0} E(Y \log Y - (1 + \log r)Y + r)$
- (vi) The function $Ent(.)$ is convex
- (vii) If $Y \leq 1$ a.s., then $Ent(e^{\lambda X}) \leq EX^2 e^{\lambda}(e^{-\lambda} + 1 + \lambda)$ for all $\lambda > 0$

The entropy

The *entropy* of a $Y > 0$ a.s. is
$Ent(Y) = EY \log Y - EY \log(EY)$. It satisfies:

- (i) $Ent(Y) \geq 0$
- (ii) $Ent(\gamma Y) = \gamma Ent(Y)$ for any $\gamma \in \mathbb{R}$
- (iii) $Ent(Y) = \sup\{EYU : Ee^U = 1\}$
- (iv) $Ent(Y) = \sup\{EY \log(V/EV) : V > 0\}$
- (v) $Ent(Y) = \inf_{r>0} E(Y \log Y - (1 + \log r)Y + r)$
- (vi) The function $Ent(.)$ is convex
- (vii) If $Y \leq 1$ a.s., then $Ent(e^{\lambda X}) \leq EX^2 e^{\lambda}(e^{-\lambda} + 1 + \lambda)$ for all $\lambda > 0$
- (viii) $Ent(Y) \leq Ent(Z) + E(Y - Z) \log(Y/EY)$ for every random variable $Z > 0$

Tensorization of entropy

The *entropy* of a $Y > 0$ a.s. is
$Ent(Y) = EY \log Y - EY \log(EY)$.

Proposition

Let T a measurable transformation of independent variables $X_1, \ldots, X_n$, then $Ent(T) \leq \sum_{i=1}^{n} EEnt_i(T)$, with $Ent_i(T) = E_i T \log T - E_i T \log E_i T$ where $E_i Y = E(Y | X_j : j \neq i)$

Entropy method

Proposition

Let T a measurable function of independent variables $X_1, ..., X_n$. Suppose there exist nonnegative constants u, σ^2 and random variables S_i and T_i for $i = 1, ..., n$ such that

Entropy method

Proposition

Let T a measurable function of independent variables $X_1, ..., X_n$. Suppose there exist nonnegative constants u, σ^2 and random variables S_i and T_i for $i = 1, ..., n$ such that

- T_i *is a measurable function of* $X_j : j \neq i$
- $S_i \leq T - T_i \leq 1$
- $E_i S_i \geq 0$
- $S_i \leq u$
- $\sum_{i=1}^{n} (T - T_i) \leq T$
- $\sum_{i=1}^{n} E_i S_i^2 \leq \sigma^2$

Entropy method

Proposition

Let T a measurable function of independent variables $X_1, \dots, X_n$. Suppose there exist nonnegative constants u, σ^2 and random variables S_i and T_i for $i = 1, \dots, n$ such that

- T_i *is a measurable function of* $X_j : j \neq i$
- $S_i \leq T - T_i \leq 1$
- $E_i S_i \geq 0$
- $S_i \leq u$
- $\sum_{i=1}^n (T - T_i) \leq T$
- $\sum_{i=1}^n E_i S_i^2 \leq \sigma^2$

then for all $t \geq 0$, we have
$P(T - ET \geq t) \leq \exp(-vh(1 + t/v))$, *where*
$h(y) = y(\log y - 1) + t$ *and* $v = (1 + u)ET + \sigma^2$.

Entropy method

Main idea of the proof: Use the results on the entropy

Step 1:

Entropy method

Main idea of the proof: Use the results on the entropy

Step 1:

- $Ent(e^{\lambda T}) \leq E\sum_{i=1}^{n} Ent_i(e^{\lambda T}) = E\sum_{i=1}^{n} e^{\lambda T_i} Ent_i(e^{\lambda(T-T_i)})$ by (ii)

Entropy method

Main idea of the proof: Use the results on the entropy

Step 1:

- $Ent(e^{\lambda T}) \leq E\sum_{i=1}^{n} Ent_i(e^{\lambda T}) = E\sum_{i=1}^{n} e^{\lambda T_i} Ent_i(e^{\lambda(T-T_i)})$ by (ii)
- For all $i = 1, ..., n$, we have $Ent_i(e^{\lambda(T-T_i)}) \leq E_i(1 - e^{\lambda(T-T_i)} + \lambda(T-T_i)e^{\lambda(T-T_i)})$ by (iv) with $r = 1$

Entropy method

Main idea of the proof: Use the results on the entropy

Step 1:

- $Ent(e^{\lambda T}) \leq E\sum_{i=1}^{n} Ent_i(e^{\lambda T}) = E\sum_{i=1}^{n} e^{\lambda T_i} Ent_i(e^{\lambda(T-T_i)})$ by (ii)
- For all $i = 1, ..., n$, we have $Ent_i(e^{\lambda(T-T_i)}) \leq E_i(1 - e^{\lambda(T-T_i)} + \lambda(T-T_i)e^{\lambda(T-T_i)})$ by (iv) with $r = 1$
- $Ent(e^{\lambda T}) \leq E\sum_{i=1}^{n}(e^{\lambda T_i} - e^{\lambda T} + \lambda(T-T_i)e^{\lambda T})$ by (ii)

Entropy method

Step 2:
Let $g_{\lambda,\alpha}(t) = e^{\lambda t}(e^{-\lambda t} - 1 + \lambda t)/(e^{\lambda t} - 1 - \lambda t + \lambda \alpha t^2)$ increasing on $(-\infty, 1]$, with $\alpha = 1/(1+u)$

Entropy method

Step 2:
Let $g_{\lambda,\alpha}(t) = e^{\lambda t}(e^{-\lambda t} - 1 + \lambda t)/(e^{\lambda t} - 1 - \lambda t + \lambda \alpha t^2)$ increasing on $(-\infty, 1]$, with $\alpha = 1/(1+u)$

- $e^{\lambda T_i} - e^{\lambda T} + \lambda(T - T_i) = g_{\lambda,\alpha}(T - T_i)(e^{\lambda T} - e^{\lambda T_i} - \lambda(T - T_i)e^{\lambda T_i} + \lambda\alpha(T - T_i)^2 e^{\lambda T_i})$

Entropy method

Step 2:
Let $g_{\lambda,\alpha}(t) = e^{\lambda t}(e^{-\lambda t} - 1 + \lambda t)/(e^{\lambda t} - 1 - \lambda t + \lambda\alpha t^2)$ increasing on $(-\infty, 1]$, with $\alpha = 1/(1+u)$

- $e^{\lambda T_i} - e^{\lambda T} + \lambda(T - T_i) = g_{\lambda,\alpha}(T - T_i)(e^{\lambda T} - e^{\lambda T_i} - \lambda(T - T_i)e^{\lambda T_i} + \lambda\alpha(T - T_i)^2 e^{\lambda T_i})$
- $e^{\lambda T_i} - e^{\lambda T} + \lambda(T - T_i) \leq g_{\lambda,\alpha}(1)(e^{\lambda T} - e^{\lambda T_i} - \lambda S_i e^{\lambda T_i} + \lambda\alpha S_i^2 e^{\lambda T_i})$

Entropy method

Step 2:

Let $g_{\lambda,\alpha}(t) = e^{\lambda t}(e^{-\lambda t} - 1 + \lambda t)/(e^{\lambda t} - 1 - \lambda t + \lambda\alpha t^2)$ increasing on $(-\infty, 1]$, with $\alpha = 1/(1+u)$

- $e^{\lambda T_i} - e^{\lambda T} + \lambda(T - T_i) = g_{\lambda,\alpha}(T-T_i)(e^{\lambda T} - e^{\lambda T_i} - \lambda(T-T_i)e^{\lambda T_i} + \lambda\alpha(T-T_i)^2 e^{\lambda T_i})$
- $e^{\lambda T_i} - e^{\lambda T} + \lambda(T - T_i) \leq g_{\lambda,\alpha}(1)(e^{\lambda T} - e^{\lambda T_i} - \lambda S_i e^{\lambda T_i} + \lambda\alpha S_i^2 e^{\lambda T_i})$
- $Ent(e^{\lambda T}) \leq g_{\lambda,\alpha}(1)E\sum_{i=1}^n(e^{\lambda T} - e^{\lambda T_i} - \lambda S_i e^{\lambda T_i} + \lambda\alpha S_i^2 e^{\lambda T_i}) \leq g_{\lambda,\alpha}(1)E\sum_{i=1}^n(e^{\lambda T} - e^{\lambda T_i} + \lambda\alpha\sigma^2 e^{\lambda T})$

Entropy method

Now, we are (supposed to notice something)

Entropy method

Now, we are (supposed to notice something)

$$Ent(e^{\lambda T}) \leq E \sum_{i=1}^{n} (e^{\lambda T_i} - e^{\lambda T} + \lambda(T - T_i)e^{\lambda T})$$

$$Ent(e^{\lambda T})/g_{\lambda,\alpha}(1) \leq E \sum_{i=1}^{n} (e^{\lambda T} - e^{\lambda T_i} + \lambda\alpha\sigma^2 e^{\lambda T})$$

Entropy method

Now, we are (supposed to notice something)

$$Ent(e^{\lambda T}) \leq E \sum_{i=1}^{n} (e^{\lambda T_i} - e^{\lambda T} + \lambda(T - T_i)e^{\lambda T})$$

$$Ent(e^{\lambda T})/g_{\lambda,\alpha}(1) \leq E \sum_{i=1}^{n} (e^{\lambda T} - e^{\lambda T_i} + \lambda\alpha\sigma^2 e^{\lambda T})$$

$$Ent(e^{\lambda T})(1 + 1/g_{\lambda,\alpha}(1)) \leq$$

$$\lambda\alpha\sigma^2 E e^{\lambda T} + E \sum_{i=1}^{n} \lambda(T - T_i)e^{\lambda T}$$

Entropy method

Starting with this,

$$Ent(e^{\lambda T})(1 + 1/g_{\lambda,\alpha}(1)) \leq$$

$$\lambda\alpha\sigma^2 Ee^{\lambda T} + E\sum_{i=1}^{n} \lambda(T - T_i)e^{\lambda T}$$

Entropy method

Starting with this,

$$Ent(e^{\lambda T})(1 + 1/g_{\lambda,\alpha}(1)) \leq$$
$$\lambda\alpha\sigma^2 E e^{\lambda T} + E\sum_{i=1}^{n} \lambda(T - T_i)e^{\lambda T}$$

we can obtain this,

$$Ent(e^{\lambda(T-ET)})(1 + 1/g_{\lambda,\alpha}(1)) \leq$$
$$\lambda\alpha v E e^{\lambda(T-ET)} + E\lambda(T - ET)e^{\lambda(T-ET)}$$

Entropy method

Step 3:
Let $F(\lambda) = Ee^{\lambda(T-ET)}$ the Laplace transform of $T - ET$
Rewriting the relevant expectations

Entropy method

Step 3:
Let $F(\lambda) = Ee^{\lambda(T-ET)}$ the Laplace transform of $T - ET$
Rewriting the relevant expectations

$$Ent(e^{\lambda(T-ET)}) = \lambda F'(\lambda) - F(\lambda)\log F(\lambda)$$
$$E(T-ET)e^{\lambda(T-ET)} = F'(\lambda)$$

Entropy method

Step 3:
Let $F(\lambda) = Ee^{\lambda(T-ET)}$ the Laplace transform of $T - ET$
Rewriting the relevant expectations

$$Ent(e^{\lambda(T-ET)}) = \lambda F'(\lambda) - F(\lambda)\log F(\lambda)$$
$$E(T-ET)e^{\lambda(T-ET)} = F'(\lambda)$$

The previous inequality becomes

$$(\log F)'(\lambda) - \log F(\lambda)\Big(\frac{e^{\lambda}-1+\alpha}{e^{\lambda}-1-\lambda+\lambda\alpha}\Big) \leq \Big(\frac{e^{\lambda}(e^{-\lambda}-1+\lambda)}{e^{\lambda}-1-\lambda+\lambda\alpha}\Big)v\alpha.$$

Entropy method

Some algebra and differential inequality leads to

$$F(\lambda) \leq v(e^{\lambda} - 1 - \lambda)$$

Entropy method

Some algebra and differential inequality leads to

$$F(\lambda) \leq v(e^{\lambda} - 1 - \lambda)$$

Step 4:
Using Markov's inequality,

$$P(T - ET \geq t) \leq e^{-\lambda t} F(\lambda) \leq \exp(-\lambda t - v(e^{\lambda} - 1 - \lambda))$$

Entropy method

Some algebra and differential inequality leads to

$$F(\lambda) \leq v(e^{\lambda} - 1 - \lambda)$$

Step 4:

Using Markov's inequality,

$$P(T - ET \geq t) \leq e^{-\lambda t} F(\lambda) \leq \exp(-\lambda t - v(e^{\lambda} - 1 - \lambda))$$

The optimal choice for λ is $\log(1 + t/v)$, which leads to the result

Throwback

Proposition

Let T a measurable function of independent variables $X_1, \dots, X_n$. Suppose there exist nonnegative constants u, σ^2 and random variables S_i and T_i for $i = 1, \dots, n$ such that

- T_i *is a measurable function of* $X_j : j \neq i$
- $S_i \leq T - T_i \leq 1$
- $E_i S_i \geq 0$
- $S_i \leq u$
- $\sum_{i=1}^n (T - T_i) \leq T$
- $\sum_{i=1}^n E_i S_i^2 \leq \sigma^2$

then for all $t \geq 0$*, we have*
$P(T - ET \geq t) \leq \exp(-vh(1 + t/v))$*, where*
$h(y) = y(\log y - 1) + t$ *and* $v = (1 + u)ET + \sigma^2$*.*

Talagrand's Inequality

We'll use this proposition to prove a stronger theorem,

Talagrand's Inequality

We'll use this proposition to prove a stronger theorem, depending on $\sigma_{\mathcal{F}}^2 = \|P(f - Pf)^2\|_{\mathcal{F}}$ the maximal variance of $\mathcal{F}$.

Talagrand's Inequality

We'll use this proposition to prove a stronger theorem, depending on $\sigma_{\mathcal{F}}^2 = \|P(f-Pf)^2\|_{\mathcal{F}}$ the maximal variance of $\mathcal{F}$.

Theorem (Bousquet-Talagrand)

For any countable class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \to \mathbb{R}$ with $f(x) - Pf \leq 1$ for every $x \in \mathcal{X}$ and $w_n := \sigma_{\mathcal{F}}^2 + 2n^{-1/2} E \sup_f \mathbb{G}_n f$ finite, then for every $t \geq 0$,

$$P\left(\sup_f \mathbb{G}_n f - E \sup_f \mathbb{G}_n f \geq t\right) \leq \exp\left(-nw_n h(1 + \frac{t}{w_n\sqrt{n}})\right)$$

where $h(y) = y(\log y - 1) + 1$.

We'll use this proposition to prove a stronger theorem, depending on $\sigma_{\mathcal{F}}^2 = \|P(f - Pf)^2\|_{\mathcal{F}}$ the maximal variance of $\mathcal{F}$.

Theorem (Bousquet-Talagrand)

For any countable class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \to \mathbb{R}$ with $f(x) - Pf \leq 1$ for every $x \in \mathcal{X}$ and $w_n := \sigma_{\mathcal{F}}^2 + 2n^{-1/2} E \sup_f \mathbb{G}_n f$ finite, then for every $t \geq 0$,

$$P\Big(\sup_f \mathbb{G}_n f - E \sup_f \mathbb{G}_n f \geq t\Big) \leq \exp\Big(-nw_n h\big(1 + \frac{t}{w_n\sqrt{n}}\big)\Big)$$

where $h(y) = y(\log y - 1) + 1$.
If $|f(x) - Pf| \leq 1$ for every $x \in \mathcal{X}$, then we can replace $\sup_f \mathbb{G}_n f$ by $\|\mathbb{G}_n\|_{\mathcal{F}}$

Proof of the Theorem

Assume w.l.g $Pf = 0$ for every f

Proof of the Theorem

Assume w.l.g $Pf = 0$ for every f
Now, let's define

- $T = \sup_{f\in\mathcal{F}} \sum_{i=1}^{n} f(X_i)$
- $T_i = \sup_{f\in\mathcal{F}} \sum_{j\neq i} f(X_j)$
- f_0 the func. involved in the definition of T for given X_i
- f_i the func. involved in the definition of T_i for given $X_j : j \neq i$
- $S_i = f_i(X_i)$

Proof of the Theorem

- T_i is a measurable function of $X_j : j \neq i$

Proof of the Theorem

- T_i is a measurable function of $X_j : j \neq i$
- $S_i \leq T - T_i \leq 1$
  - $S_i + T_i = \sum_{j=1}^{n} f_i(X_j) \leq T$
  - $T - T_i \leq 1$

Proof of the Theorem

- T_i is a measurable function of $X_j : j \neq i$
- $S_i \leq T - T_i \leq 1$
  - $S_i + T_i = \sum_{j=1}^{n} f_i(X_j) \leq T$
  - $T - T_i \leq 1$
- $E_i S_i = Pf_i = 0$

Proof of the Theorem

- T_i is a measurable function of $X_j : j \neq i$
- $S_i \leq T - T_i \leq 1$
  - $S_i + T_i = \sum_{j=1}^{n} f_i(X_j) \leq T$
  - $T - T_i \leq 1$
- $E_i S_i = P f_i = 0$
- $\sum_{i=1}^{n} E_i S_i^2 = \sum_{i=1}^{n} P f_i^2 \leq n\sigma_{\mathcal{F}}^2$

Proof of the Theorem

- T_i is a measurable function of $X_j : j \neq i$
- $S_i \leq T - T_i \leq 1$
    - $S_i + T_i = \sum_{j=1}^{n} f_i(X_j) \leq T$
    - $T - T_i \leq 1$
- $E_i S_i = Pf_i = 0$
- $\sum_{i=1}^{n} E_i S_i^2 = \sum_{i=1}^{n} Pf_i^2 \leq n\sigma_{\mathcal{F}}^2$
- $\sum_{i=1}^{n}(T - T_i) \leq T$
    - $(n-1)T = \sum_{i=1}^{n} \sum_{j \neq i} f_0(X - j)$
    - $\sum_{i=1}^{n} \sum_{j \neq i} f_0(X - j) \leq \sum_{i=1}^{n} T_i$
    - $\sum_{i=1}^{n} T \leq T + \sum_{i=1}^{n} T_i$

More complete theorem

Theorem (Bousquet-Talagrand)

For any countable class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \to \mathbb{R}$ with $|f(x) - Pf| \leq 1$ for every $x \in \mathcal{X}$ and $w_n := \sigma^2_{\mathcal{F}} + 2n^{-1/2} E \sup_f \mathbb{G}_n f$ finite, then for every $t \geq 0$,

$$P\Big(\sup_f \mathbb{G}_n f - E \sup_f \mathbb{G}_n f \geq t\Big) \leq \exp\Big(-n w_n h(1 + \frac{t}{w_n\sqrt{n}})\Big)$$

$$P\Big(\sup_f \mathbb{G}_n f - E \sup_f \mathbb{G}_n f \leq -t\Big) \leq \exp\Big(-n \frac{w_n}{9} h(1 + \frac{t}{w_n\sqrt{n}})\Big)$$

where $h(y) = y(\log y - 1) + 1$.

More complete theorem

Theorem (Bousquet-Talagrand)

For any countable class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \to \mathbb{R}$ with $|f(x) - Pf| \leq 1$ for every $x \in \mathcal{X}$ and $w_n := \sigma_{\mathcal{F}}^2 + 2n^{-1/2} E \sup_f \mathbb{G}_n f$ finite, then for every $t \geq 0$,

$$P\Big(\sup_f \mathbb{G}_n f - E \sup_f \mathbb{G}_n f \geq t\Big) \leq \exp\Big(-n w_n h(1 + \frac{t}{w_n \sqrt{n}})\Big)$$

$$P\Big(\sup_f \mathbb{G}_n f - E \sup_f \mathbb{G}_n f \leq -t\Big) \leq \exp\Big(-n \frac{w_n}{9} h(1 + \frac{t}{w_n \sqrt{n}})\Big)$$

where $h(y) = y(\log y - 1) + 1$.

The proof of the second inequality is also based on the entropy method

Corollary

Corollary (Bousquet-Talagrand)

For any countable class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \to \mathbb{R}$ with $\|f\|_\infty M$ and $Pf^2 \leq \delta^2$ for every $f \in \mathcal{F}$. Then for every $t > 0$,

$$\forall g \in \mathcal{F} : \frac{1}{2}\mathbb{G}_n g \leq E \sup_{f \in \mathcal{F}} \mathbb{G}_n f + \frac{Mt}{\sqrt{n}} + \delta\sqrt{t}$$

with probability at least $1 - e^{-t}$.

Thank you

Any questions?