Symmetrization and Measurability

Wenyi Xie

University of North Carolina at Chapel Hill

July 14, 2021

1. Application of General Maximal Inequality
2. Symmetrization Inequality
3. Measurability

From previous talk

Orlicz norm

For a nondecreasing, nonzero convex function $\psi : [0, \infty] \mapsto [0, \infty]$, with $\psi(0) = 0$, the Orlicz norm $\|X\|_\psi$ of a real random variable X is defined as

$$\|X\|_\psi \equiv \inf \left\{ c > 0 : \mathrm{E}\psi\left(\frac{|X|}{c}\right) \leq 1 \right\}$$

Theorem 8.4 (General maximal inequality)

Let ψ satisfy the conditions of Lemma 8.2, and let $\{X(t), t \in T\}$ be a separable stochastic process with $\|X(s) - X(t)\|_\psi \leq r d(s,t)$, for all $s, t \in T$. Then for any $\eta, \delta > 0$,

$$\left\| \sup_{s,t \in T : d(s,t) \leq \delta} |X(s) - X(t)| \right\|_\psi \leq K \left[\int_0^\eta \psi^{-1}(D(\epsilon, d))\, d\epsilon + \delta \psi^{-1}\left(D^2(\eta, d)\right) \right]$$

An important application of previous theorem is to sub-Gaussian processes

Corollary 8.5

Let $\{X(t), t \in T\}$ be a separable sub-Gaussian process with respect to d. Then for all $\delta > 0$,

$$\mathrm{E}\left(\sup_{s,t\in T: d(s,t)\leq\delta} |X(s) - X(t)|\right) \leq K \int_0^{\delta} \sqrt{\log D(\epsilon, d)}\, d\epsilon$$

where K is a universal constant. Also, for any $t_0 \in T$,

$$\mathrm{E}\left(\sup_{t\in T} |X(t)|\right) \leq \mathrm{E}\,|X(t_0)| + K \int_0^{\mathrm{diam}\ T} \sqrt{\log D(\epsilon, d)}\, d\epsilon$$

Proof for Corollary 8.5

$$\psi_p(x) \equiv e^{x^p} - 1$$

Based on Exercise 8.5.4, ψ_p satisfies the conditions of Lemma 8.2 with $c = 1$, for any $p \in [1, \infty]$

First apply Theorem 8.4 with $\psi = \psi_2$ and $\eta = \delta$, we have

$$\left\| \sup_{s,t \in T: d(s,t) \leq \delta} |X(s) - X(t)| \right\|_{\psi_2} \leq K \left[\int_0^\delta \psi_2^{-1}(D(\epsilon, d))\, d\epsilon + \delta \psi_2^{-1}\left(D^2(\delta, d)\right) \right]$$

$$\psi_2^{-1}\left(D^2(\delta, d)\right) = \sqrt{\log(1 + D^2(\delta, d))} \leq \sqrt{2 \log(1 + D(\delta, d))} = \sqrt{2}\psi_2^{-1}\left(D(\delta, d)\right)$$

The second term in the first inequality can be replaced by

$$\sqrt{2}\delta\psi_2^{-1}(D(\delta, d)) \leq \sqrt{2} \int_0^\delta \psi_2^{-1}(D(\epsilon, d))\, d\epsilon$$

and we obtain

$$\left\| \sup_{d(s,t)\leq\delta} |X(s) - X(t)| \right\|_{\psi_2} \leq K \int_0^\delta \sqrt{\log(1 + D(\epsilon, d))}\, d\epsilon$$

Note that $D(\epsilon, d) \geq 2$ for all ϵ strictly less than the diam T. And $(1 + m) \leq m^2$ for all $m \geq 2$, the 1 inside of the logarithm can be further removed at the cost of increasing K again, whenever $\delta < \operatorname{diam} T$.

For ψ_2, $E[|X|] \leq \|X\|_{\psi_2}$, this completes the first part of the corollary.

And the second statement of Corollary 8.5 is simply an easy consequence of the first, the proof is complete.

Rademacher Process

Now consider an important sub-Gaussian process: the Rademacher Process

$$X(a) = \sum_{i=1}^{n} \epsilon_i a_i, \quad a \in \mathbb{R}^n$$

where $\epsilon_1, \dots, \epsilon_n$ are i.i.d Rademacher random variables satisfying
$P(\epsilon = -1) = P(\epsilon = 1) = 1/2$
This process will emerge in our development of Donsker results based on uniform entropy.
The following lemma verifies that Rademacher processes are sub-Gaussian

Lemma 8.7 (Hoeffding's inequality)

Let $a = (a_1, \cdots, a_n) \in \mathcal{R}^n$ and $\epsilon_1, \dots, \epsilon_n$ be independent Rademacher random variables. Then

$$\mathrm{pr}\left(\left|\sum_{i=1}^{n} \epsilon_i a_i\right| > x\right) \leq 2e^{-\frac{1}{2}x^2/\|a\|^2}$$

for the Euclidean norm $\|\cdot\|$

Proof for Lemma 8.7

For any λ and Rademacher variable ϵ, one has

$$\mathrm{E} e^{\lambda\epsilon} = \left(e^{\lambda} + e^{-\lambda}\right)/2 = \sum_{i=0}^{\infty} \lambda^{2i}/(2i)! \leq e^{\lambda^2/2}$$

where the last inequality follows from the relationship that $(2i)! \geq 2^i i!$ for all nonnegative integers.
Hence Markov's inequality gives for any $\lambda > 0$

$$\mathrm{pr}\left(\sum_{i=1}^{n} \epsilon_i a_i > x\right) \leq e^{-\lambda x} \mathrm{E}\exp\left\{\lambda \sum_{i=1}^{n} \epsilon_i a_i\right\} \leq \exp\left\{\left(\lambda^2/2\right)\|a\|^2 - \lambda x\right\}$$

Setting $\lambda = x/\|a\|^2$ yields the desired upper bound.
Since multiplying the Rademacher random variables by -1 does not change the joint distribution, we have

$$\mathrm{P}\left(-\sum_{i=1}^{n} \epsilon_i a_i > x\right) = \mathrm{P}\left(\sum_{i=1}^{n} \epsilon_i a_i > x\right)$$

The hoeffding's inequality shows that

$$\mathrm{pr}\left(\left|\sum_{i=1}^{n}\epsilon_i a_i\right| > x\right) \le 2e^{-\frac{1}{2}x^2/\|a\|^2}$$

We also have $\|\sum \epsilon a\|_{\psi_2} \le \sqrt{6}\|a\|$ This is because by Lemma 8.1,

$$\mathrm{P}(|\sum_{i=1}^{n}\epsilon_i a_i| > x) \le Ke^{-Cx^p}, \text{ for all } x > 0$$

$$\Rightarrow \|X\|_{\psi_p} \le ((1+K)/C)^{1/p}$$

In this case, $K = 2, p = 2, C = 1/(2\|a\|^2)$

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

The Symmetrization

One of the two main approaches toward deriving GC/Donsker theorem is based on comparing the empircal process to a symmetrized empirical process.
Let $\epsilon_1, \ldots, \epsilon_n$ be independent Rademacher random variables, which are also independent of X_i
Instead of the empirical process

$$f \mapsto (\mathbb{P}_n - P) f = \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - Pf)$$

consider the symmetrized process

$$f \mapsto \mathbb{P}_n^{\circ} f \equiv n^{-1} \sum_{i}^{n} \epsilon_i f(X_i)$$

Note that both processes have mean function zero.

And G-C theorem or Donsker theorem holds for one of these processes if and only if it holds for the other.

The goal is to pass from $\mathbb{P}_n - P$ to $\mathbb{P}_n^\circ$ and apply arguments conditionally on the original $X's$. Then for fixed X, the symmetrized empirical measure is a Rademacher process, hence a sub-Gaussian process, then Corollary 8.5 can be applied.

Before moving onto bound the maxima and moduli of the process $\mathbb{P}_n - P$, one needs to be careful about the possible nonmeasurability of the suprema of $\|\mathbb{P}_n - P\|_{\mathcal{F}}$

The results will be formulated in terms of outer expectation and we need to be clear about the underlying product probability space so that outer expectations are well-defined.

Throughout this section, if outer expectation is involved, we will assume that $X_1, \cdots, X_n$ are the coordinate projection of the product space $(\mathcal{X}^n, \mathcal{A}^n, P^n)$, where $\mathcal{A}^n$ is the product σ-field generated from the sets $A_1 \times \cdots \times A_n$ for $A_1, \cdots, A_n \in \mathcal{A}$ and the outer expectation of functions $(X_1, \ldots, X_n) \mapsto h(X_1, \ldots, X_n)$ are computed for P^n.

"Independent" here is understood in terms of a product probability space. With anxilliary variables independent of $X's$ Z, the underlying probability space is assumed to be of the form $(\mathcal{X}^n, \mathcal{A}^n, P^n) \times (\mathcal{Z}, \mathcal{C}, Q)$, with X_i equal to the coordinate projections on the i^{th} coordinate and Z depend only on the $(n+1)^{th}$ coordinate.

Theorem 8.8 (Symmetrization)

For every nondecreasing, convex $\phi : \mathbb{R} \mapsto \mathbb{R}$ and class of measurable functions $\mathcal{F}$,

$$\begin{aligned} \mathrm{E}^* \phi\left(\frac{1}{2}\left\|\mathbb{P}_n - P\right\|_{\mathcal{F}}\right) &\leq \mathrm{E}^* \phi\left(\left\|\mathbb{P}_n^{\circ}\right\|_{\mathcal{F}}\right) \\ &\leq \mathrm{E}^* \phi\left(2\left\|\mathbb{P}_n - P\right\|_{\mathcal{F}} + |R_n| \cdot \|P\|_{\mathcal{F}}\right) \end{aligned}$$

where $R_n \equiv \mathbb{P}_n^{\circ} 1 = n^{-1} \sum_{i=1}^{n} \epsilon_i$ and the outer expectation are computed based on the product σ-field described above.

This lemma will mostly be used with the choice $\phi(x) = x$

Proof for Theorem 8.8

Let $Y_1, \cdots, Y_n$ be independent copies of $X_1, \cdots, X_n$, defined as the coordinate projections on the last n coordinates in the product space
$(\mathcal{X}^n, \mathcal{A}^n, P^n) \times (\mathcal{Z}, \mathcal{C}, Q) \times (\mathcal{X}^n, \mathcal{A}^n, P^n)$
$(\mathcal{Z}, \mathcal{C}, Q)$ is the probability space for the vector of independent Rademacher random variables involved in $\mathbb{P}_n^\circ$

For fixed values $X_1, \cdots, X_n$,

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} \frac{1}{n}\left|\sum_{i=1}^{n}[f(X_i) - \mathrm{E}f(Y_i)]\right|$$
$$\leq \mathrm{E}^*_Y \sup_{f\in\mathcal{F}} \frac{1}{n}\left|\sum_{i=1}^{n}[f(X_i) - f(Y_i)]\right|$$

E^*_Y is the outer expectation with respect to Y and by treating X as constants and using the probability space $(\mathcal{X}^n, \mathcal{A}^n, P^n)$
And Jensen's inequality yields,

$$\Phi\left(\|\mathbb{P}_n - P\|_{\mathcal{F}}\right) \leq \mathrm{E}_Y \Phi\left(\left\|\frac{1}{n}\sum_{i=1}^{n}[f(X_i) - f(Y_i)]\right\|_{\mathcal{F}}^{*Y}\right)$$

$*Y$ denotes the minimal measurable majorant of the supremum, still with X being fixed.

Because $\phi(x)$ is nondecreasing and continuous, the $*Y$ inside ϕ can be moved to E_Y^*

$$\Phi\left(\|\mathbb{P}_n - P\|_{\mathcal{F}}\right) \leq \mathrm{E}_Y^*\Phi\left(\left\|\frac{1}{n}\sum_{i=1}^{n}[f(X_i) - f(Y_i)]\right\|_{\mathcal{F}}\right)$$

Next take expectation with respect to X,

$$\begin{aligned} \mathrm{E}^*\Phi\left(\|\mathbb{P}_n - P\|_{\mathcal{F}}\right) &\leq \mathrm{E}_X^*\mathrm{E}_Y^*\Phi\left(\frac{1}{n}\left\|\sum_{i=1}^{n}[f(X_i) - f(Y_i)]\right\|_{\mathcal{F}}\right) \\ &\leq \mathrm{E}^*\Phi\left(\frac{1}{n}\left\|\sum_{i=1}^{n}[f(X_i) - f(Y_i)]\right\|_{\mathcal{F}}\right) \end{aligned}$$

The second inequality holds since double outer expectation can be bounded above by the joint outer expectation by the Fubini's theorem for outer expectation (Lemma 6.14)

Next we establish the connection with symmetrized process.
Note that adding a minus sign in front of a term $f(X_i) - f(Y_i)$ is the same as exchanging X_i and Y_i. And by constructing an underlying product probability space, the outer expectation of any function $f(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ remains unchanged under permutation.
For any n-vector $(e_1, \ldots, e_n) \in \{-1, 1\}^n$, $\left\| n^{-1} \sum_{i=1}^n e_i \left[f(X_i) - f(Y_i)\right] \right\|_{\mathcal{F}}$ is just a permutation of

$$h(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \equiv \left\| n^{-1} \sum_{i=1}^n \left[f(X_i) - f(Y_i)\right] \right\|_{\mathcal{F}}$$

Expression from last step:

$$\begin{aligned} \Rightarrow \mathrm{E}^* \Phi\left(\|\mathbb{P}_n - P\|_{\mathcal{F}}\right) &\leq \mathrm{E}^* \Phi\left(\frac{1}{n} \left\| \sum_{i=1}^n \left[f(X_i) - f(Y_i)\right] \right\|_{\mathcal{F}} \right) \\ &= \mathrm{E}_\epsilon \mathrm{E}^*_{X,Y} \Phi\left(\left\| n^{-1} \sum_{i=1}^n e_i \left[f(X_i) - f(Y_i)\right] \right\|_{\mathcal{F}} \right) \end{aligned}$$

Next use the triangle inequality to separate the contributions of $X's$ and $Y's$

$$\mathrm{E}_\epsilon \mathrm{E}^*_{X,Y} \Phi\left(\left\| n^{-1} \sum_{i=1}^n e_i \left[f(X_i) - f(Y_i)\right]\right\|_{\mathcal{F}}\right)$$

can be bounded by

$$\frac{1}{2}\mathrm{E}_\varepsilon \mathrm{E}^*_{X,Y} \Phi\left(2\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathcal{F}}\right) + \frac{1}{2}\mathrm{E}_\varepsilon \mathrm{E}^*_{X,Y} \Phi\left(2\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(Y_i)\right\|_{\mathcal{F}}\right)$$

The double outer expectation can be replaced by a joint outer expectation. Thus we have proven the first statement:

$$\mathrm{E}^*\Phi\left(\|\mathbb{P}_n - P\|_{\mathcal{F}}\right) \leq \mathrm{E}^*\Phi\left(2\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(Y_i)\right\|_{\mathcal{F}}\right)$$

For the second inequality in Theorem 8.8 that,

$$\mathrm{E}^*\phi\left(\|\mathbb{P}_n^\circ\|_{\mathcal{F}}\right) \leq \mathrm{E}^*\phi\left(2\left\|\mathbb{P}_n - P\right\|_{\mathcal{F}} + |R_n| \cdot \|P\|_{\mathcal{F}}\right)$$

Still define Y as independent copied of $X_1, \cdots, X_n$ as before. Holding $X_1, \cdots, X_n$ and $\epsilon_1, \cdots, \epsilon_n$ fixed, we have

$$\begin{aligned} \|\mathbb{P}_n^\circ f\|_{\mathcal{F}} &= \|\mathbb{P}_n^\circ (f - Pf) + \mathbb{P}_n^\circ Pf\|_{\mathcal{F}} \\ &= \|\mathbb{P}_n^\circ (f - Ef(Y)) + R_n Pf\|_{\mathcal{F}} \\ &\leq E_Y^* \left\| \frac{1}{n}\sum_{i=1}^n \epsilon_i \left[f(X_i) - f(Y_i)\right]\right\|_{\mathcal{F}} + |R_n| \cdot \|P\|_{\mathcal{F}} \end{aligned}$$

Apply Jensen's inequality,

$$\Phi\left(\|\mathbb{P}_n^\circ\|_{\mathcal{F}}\right) \leq \mathrm{E}_Y^*\Phi\left(\left\| \frac{1}{n}\sum_{i=1}^n \epsilon_i \left[f(X_i) - f(Y_i)\right]\right\|_{\mathcal{F}} + |R_n| \cdot \|P\|_{\mathcal{F}}\right)$$

Then we take outer expectation on both sides with respect to X and ϵ. With the same permutation trick we have used in previous proof, ϵ can be replaced by 1's. This gives us

$$\mathrm{E}^*\Phi\left(\|\mathbb{P}_n^\circ\|_{\mathcal{F}}\right) \leq \mathrm{E}_\epsilon \mathrm{E}_X^* \mathrm{E}_Y^* \Phi\left(\left\|\frac{1}{n}\sum_{i=1}^n [f(X_i) - f(Y_i)]\right\|_{\mathcal{F}} + |R_n| \cdot \|P\|_{\mathcal{F}}\right)$$

And by adding and subtracting $\mathbb{P}f$ in the summation and triangle inequality, we can further bound the right-hand-side by

$$\frac{1}{2}\mathrm{E}_\epsilon E_X^* \mathrm{E}_Y^* \phi\left(2\left\|\frac{1}{n}\sum_{i=1}^n [f(X_i) - Pf]\right\|_{\mathcal{F}} + |R_n| \cdot \|P\|_{\mathcal{F}}\right)$$
$$+ \frac{1}{2}\mathrm{E}_\epsilon \mathrm{E}_X^* \mathrm{E}_Y^* \phi\left(2\left\|\frac{1}{n}\sum_{i=1}^n [f(Y_i) - Pf]\right\|_{\mathcal{F}} + |R_n| \cdot \|P\|_{\mathcal{F}}\right)$$

By outer Fubini's theorem, it can be bounded by

$$\mathrm{E}^*\phi\left(2\|\mathbb{P}_n - P\|_{\mathcal{F}} + |R_n| \cdot \|P\|_{\mathcal{F}}\right)$$

Thus completes the proof.

1. Application of General Maximal Inequality
2. Symmetrization Inequality
3. Measurability

Measurability

The above symmetrization results will be most useful when the supremum $\|\mathbb{P}_n^\circ\|_{\mathcal{F}}$ is measurable and Fubini's theorem permits first taking expectation wrt $\epsilon|X$ and then X

If the supremum is not measurable, only weak Fubini's theorem applies, and the reordering of expectations may not be valid.

In this case, we assume that the class $\mathcal{F}$ is a P-measurable class.

P-measurable class

A class $\mathcal{F}$ of measurable function $f: \mathcal{X} \mapsto \mathbb{R}$, on the probability space $(\mathcal{X}, \mathcal{A}, P)$ is P-measurable if

$$(X_1, \ldots, X_n) \mapsto \left\| \sum_{i=1}^{n} e_i f(X_i) \right\|_{\mathcal{F}}$$

is measurable on the completion of $(\mathcal{X}^n, \mathcal{A}^n, P^n)$ for every n and every vector $(e_1, \ldots, e_n) \in \mathbb{R}^n$

Another assumption on $\mathcal{F}$ that is stronger than P-measurability but often easier to verify is pointwise measurability.

Pointwise measurability

A class $\mathcal{F}$ of measurable functions is pointwise measurable if there exists a countable set $\mathcal{G} \subset \mathcal{F}$ such that for every $f \in \mathcal{F}$, there exists a sequence $\{g_m\} \in \mathcal{G}$ with $g_m(x) \to f(x)$ for every $x \in \mathcal{X}$

An example of pointwise measurable,
the class $\mathcal{F} = \{\mathbf{1}\{x \leq t\} : t \in \mathbb{R}\}$ where the sample space $\mathcal{X} = \mathbb{R}$ is pointwise measurable.
Let $\mathcal{G} = \{1\{x \leq t\} : t \in \mathbb{Q}\}$, and fix the function

$$x \mapsto f(x) = \mathbf{1}\{x \leq t_0\}$$

for some $t_0 \in \mathcal{R}$
$\mathcal{G}$ is countable. Let $\{t_m\}$ be a sequence of rationals with $t_m \geq t_0$ for all $m \geq 1$ and with $t_m \downarrow t_0$.
Then $x \mapsto g_m(x) = \mathbf{1}\{x \leq t_m\}$ satisfies that $g_m \in \mathcal{G}$, and $g_m(x) \to f(x)$ for all $x \in \mathbb{R}$.
t_0 is arbituary, therefore the statement holds for all functions in $\mathcal{F}$, which shows that $\mathcal{F}$ is pointwise measurable.

Besides easy to verify, another nice feature of pointwise measurable class is that they have a number of useful preservation features.

- When $\mathcal{F}_1$ and $\mathcal{F}_2$ are pointwise measurable, so is $\mathcal{F}_1 \cup \mathcal{F}_2$
- $\mathcal{F}_1 \wedge \mathcal{F}_2$ (all possible pairwise minimums) is PM
- $\mathcal{F}_1 \vee \mathcal{F}_2$ (all possible pairwise maximum) is PM
- $\mathcal{F}_1 + \mathcal{F}_2$ is PM
- $\mathcal{F}_1 \cdot \mathcal{F}_2 \equiv \{f_1 f_2 : f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$ is PM

Lemma 8.10

Let $\mathcal{F}_1, \ldots, \mathcal{F}_k$ be PM classes of real functions on $\mathcal{X}$, and let $\phi : \mathbb{R}^k \mapsto \mathbb{R}$ be continuous.
Then the class $\phi \circ (\mathcal{F}_1, \ldots, \mathcal{F}_k)$ is PM, where $\phi \circ (\mathcal{F}_1, \ldots, \mathcal{F}_k)$ denotes the class

$$\{\phi(f_1, \ldots, f_k) : (f_1, \ldots, f_k) \in \mathcal{F}_1 \times \cdots \times \mathcal{F}_k\}$$