Maximal Inequalities and Covering Numbers

1. Why maximal inequalities? To bound probabilities involving suprema of r.v.s, for both the law of large numbers and the central limit theorem.

2. Orlicz norm $||X||_\psi$:
   For X a random variable, ψ a non-decreasing, convex function with $\psi(0) = 0$,
   $$||X||_\psi = \inf\{C > 0 : E\psi\Big(\frac{|X|}{C}\Big) \le 1\}$$
   $\inf \emptyset = \infty$

   Example: For $\psi(x) = x^p, \ p \ge 1$
   $$\begin{aligned} & E\psi\Big(\frac{|X|}{C}\Big) = E\big(\frac{|X|^p}{C^p}\big) \le 1, \qquad C > 0 \\ \Leftrightarrow \ & \big(E|X|^p\big)^{1/p} \le C \\ \Rightarrow \ & ||X||_\psi \le \big(E|X|^p\big)^{1/p} \equiv ||X||_p = \ L_p \text{ norm} \\ & E\Big(\frac{|X|}{||X||_\psi}\Big)^p \le 1 \Leftrightarrow \big(E|X|^p\big)^{1/p} \le ||X||_\psi \\ \Rightarrow \ & ||X||_\psi = ||X||_p \end{aligned}$$

Primary interest: $\psi_p(x) \equiv \exp(x^p) - 1,\ p \geq 1$

$$x^p \leq \psi_p(x) = e^{x^p} - 1, \qquad x \geq 0, \qquad \text{(by Taylor expansion)}$$

$$\Rightarrow \ E\Big(\frac{|X|}{||X||_{\psi_p}}\Big)^p \leq E\psi_p\Big(\frac{|X|}{||X||_{\psi_p}}\Big)$$

$$\Rightarrow \ ||X||_p = \big(E|X|^p\big)^{1/p} \leq \Big\{E\psi_p\Big(\frac{|X|}{||X||_{\psi_p}}\Big)\Big\}^{1/p}||X||_{\psi_p}$$

$$\leq ||X||_{\psi_p}$$

Also

$$||X||_{\psi_p} \leq ||X||_{\psi_q}(\log 2)^{1/q-1/p}, \qquad p \leq q$$
$$||X||p \leq p!||X||_{\psi_1}$$

3. Use Orlicz norm to estimate the tail of a distribution:

$$\begin{aligned} P\big(|X| > x\big) &\leq P\Big(\psi(|X|/||X||_\psi) \geq \psi(|x|/||X||_\psi)\Big) \\ &\leq E\Big(\frac{\psi(|X|/||X||_\psi)}{\psi(|x|/||X||_\psi)}\Big) \qquad \text{(Markov's inequality)} \\ &\leq \frac{1}{\psi(|x|/||X||_\psi)} \end{aligned}$$

For $\psi_p = \exp(x^p) - 1$, if $||X||_{\psi_p} < \infty$,

$$\frac{1}{\psi_p(|x|/||X||_{\psi_p})} \approx \exp(-Cx^p)$$

Conversely,

4. Lemma 2.2.1:

   For X a random variable with $P(|X| > x) \le K \exp(-Cx^p)$ for every x, for constants K and C, and for $p \ge 1$. Then

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

   Proof:

   $||X||_{\psi_p} = \inf\{A > 0 : E\Big(e^{(|X|/A)^p} - 1\Big) \le 1\}$

   Let $A = D^{-1/p}$.

$$\begin{aligned}
& E\Big(e^{D|X|^p} - 1\Big) = E\int_0^{|X|^p} De^{Ds}ds \\
= & \; E\int_0^{\infty} 1(|X|^p > s)De^{Ds}ds \\
= & \int_0^{\infty} P(|X| > s^{1/p})De^{Ds}ds \qquad \text{(Fubini's theorem)} \\
\le & \int_0^{\infty} KDe^{(D-C)s}ds \\
= & \; KD/(C-D), \qquad \text{if } C > D \\
\le & \; 1, \qquad \text{if } C/(K+1) \ge D \\
& \qquad \Leftrightarrow A = D^{-1/p} \ge \big((K+1)/C\big)^{1/p} \\
\Rightarrow & \; ||X||_{\psi_p} \le \big((K+1)/C\big)^{1/p}
\end{aligned}$$

5. ψ-norm of a maximum of finite many random variables:

Lemma 2.2.2
Let ψ be a convex, nondecreasing, nonzero function with $\psi(0) = 0$ and $\overline{\lim}_{x,y\to\infty}\psi(x)\psi(y)/\psi(cxy) < \infty$ for some constant c.
Then for any random variables $X_1, \cdots, X_m$

$$||\max_{1\le i\le m} X_i||_\psi \le K\psi^{-1}(m)\max_i ||X_i||_\psi$$

for a constant K depending only on ψ.

Proof:
For simplicity, assume $\psi(x)\psi(y) \le \psi(cxy), \forall x, y \ge 1$, so

$$\psi(x/y) \le \psi(cx)/\psi(y), \qquad \forall x \ge y \ge 1$$

Now, for $y \geq 1$ and any $C > 0$,

$$
\begin{aligned}
& \max \psi\Big(\frac{|X_i|}{Cy}\Big) \\
\leq \ & \max \Big[\frac{\psi(c|X_i|/C)}{\psi(y)} 1\Big\{\frac{|X_i|}{Cy} \geq 1\Big\} + \psi\Big(\frac{|X_i|}{Cy}\Big) 1\Big\{\frac{|X_i|}{Cy} < 1\Big\}\Big] \\
\leq \ & \max \frac{\psi(c|X_i|/C)}{\psi(y)} + \psi(1) \\
\leq \ & \sum \frac{\psi(c|X_i|/C)}{\psi(y)} + \psi(1)
\end{aligned}
$$

Set $C = c \max ||X_i||_\psi$,

$$
\begin{aligned}
E\psi\Big(\frac{\max |X_i|}{Cy}\Big) & \leq E \max \psi\Big(\frac{|X_i|}{Cy}\Big) \\
\leq \ & \sum E\Big(\frac{\psi(|X_i|/\max ||X_i||_\psi)}{\psi(y)}\Big) + \psi(1) \\
\leq \ & \frac{m}{\psi(y)} + \psi(1) \\
\leq \ & 1 \qquad \text{if } \tfrac{m}{1-\psi(1)} \leq y \text{ when } \psi(1) \leq 1/2 \\
& \Leftrightarrow y \geq \psi^{-1}(2m) \\
& \Leftrightarrow Cy \geq \psi^{-1}(2m) c \max ||X_i||_\psi
\end{aligned}
$$

Thus,

$$
\begin{aligned}
||\max_i X_i||_\psi \ & \leq \ \psi^{-1}(2m) c \max_i ||X_i||_\psi \\
& \leq \ 2c\psi^{-1}(m) \max_i ||X_i||_\psi \\
& \qquad (\psi \text{ is convex and } \psi(0) = 0)
\end{aligned}
$$

Finally, for general ψ, note $\exists \sigma \leq 1$ and $\tau > 0$ such that $\phi(x) = \sigma\psi(\tau x)$ satisfies $\phi(x)\phi(y) \leq \phi(cxy), \forall x, y \geq 1$ Apply inequality to ϕ and note $||X||_\psi \leq ||X||_\phi/(\sigma\tau) \leq ||X||_\phi/\sigma$.

6. Use *chaining method* to extended to general maximum over infinitely many variables:

   Definitions:

   Let (T, d) be arbitrary semi-metric space.
   Covering number $N(\epsilon, d)$ is the minimal number of balls of radius ϵ needed to cover T.
   A collection of points is ϵ-*separated* if the distance between each pair of points $> \epsilon$.
   Packing number $D(\epsilon, d)$ is the maximum number of ϵ-separated points in T.

$$N(\epsilon, d) \leq D(\epsilon, d) \leq N(\epsilon/2, d)$$

Theorem 2.2.4:

Let ψ be a convex, nondecreasing, nonzero function with $\psi(0) = 0$ and $\overline{\lim}_{x,y\to\infty}\psi(x)\psi(y)/\psi(cxy) < \infty$ for some constant c. Let $\{X_t : t \in T\}$ be a separable stochastic process with

$$||X_s - X_t||_\psi \leq Cd(s,t), \qquad \forall s,t$$

for some semi-metric d on T and constant C. Then, for any $\eta, \delta > 0$,

$$\Big|\Big| \sup_{d(s,t)\leq\delta} |X_s - X_t| \Big|\Big|_\psi \leq K\Big[\int_0^\eta \psi^{-1}(D(\epsilon,d))d\epsilon + \delta\psi^{-1}(D^2(\eta,d))\Big]$$

Corollary 2.2.5:

The constant K can be chosen such that

$$\Big|\Big| \sup_{s,t} |X_s - X_t| \Big|\Big|_\psi \leq K\Big[\int_0^{\text{diam } T} \psi^{-1}(D(\epsilon,d))d\epsilon\Big]$$

where diam T is the diameter of T.