20.1 Chaining and Orlicz Processes

**Definition 20.1 (χ_q Norm)** The \(\chi_q\) norm of a random variable \(X\) with mean zero is

\[
\|X\|_{\chi_q} = \inf \{ \lambda > 0 : E \left[ \frac{|X|}{\lambda} \right] \leq 1 \} \tag{20.1}
\]

where \(\chi_q(x) = e^{x^q} - 1\) for \(q \in [1, 2]\). If no such \(\lambda\) exists, \(\|X\|_{\chi_q} = \infty\).

Note that

\[
P(|X| > t) = P \left( \chi_q \left( \frac{|X|}{\|X\|_{\chi_q}} \right) > \chi_q \left[ \frac{t}{\|X\|_{\chi_q}} \right] \right)
\]

because \(\chi_q\) increasing \(\leq \frac{1}{\chi_q \left( \frac{t}{\|X\|_{\chi_q}} \right)}\) by Markov \(\tag{20.3}\)

We will show on a later homework assignment that this implies

\[
P(|X| > t) \leq c_q \exp \{-c_2 t^q\} \tag{20.4}
\]

which shows that we are simply defining a generalized notion of concentration, with Sub-Gaussian \((q = 2)\) and Sub-Exponential \((q = 1)\) tail decay as two special cases.

Further note that if \(X_1, ..., X_n\) iid w/ \(\|X_i\|_{\chi_q} = \sigma^2\) then:

\[
E \left[ \max_{i=1, ..., n} X_i \right] \leq \sigma \chi_q^{-1}(n) \tag{20.5}
\]

**Remark 1** If \(\chi(u) = u^p, p \geq 1,\) then
\[ \|X\|_\chi = \left( \mathbb{E}[|X|^p] \right)^{\frac{1}{p}} \] (20.6)

More generally, any function \( \chi : \mathbb{R}_+ \to \mathbb{R}_+ \) strictly increasing, convex and with \( \chi(0) = 0 \) would yield a norm \( \| \cdot \|_\chi \) on the space of zero-mean RV’s. We call these Orlicz norms.

We will focus on \( \chi_q(x) = e^{x^q} - 1 \) from here on in.

**Definition 20.2 (\( \chi_q \) process)** Let \( \{\mathbb{T}, \rho\} \) be a metric space. A zero-mean stochastic process \( \{X_\theta : \theta \in \mathbb{T}\} \) is a \( \chi_q \) process if

\[ \|X_\theta - X_{\theta'}\|_{\chi_q} \leq \rho(\theta, \theta') \quad \forall \theta, \theta' \in \mathbb{T} \] (20.7)

As an example, the Gaussian process \( G_\theta = \{\langle \theta, w \rangle, \theta \in \mathbb{T}\}, w \sim \mathcal{N}(0, I) \) is also a \( \chi_2 \) process, with \( \rho(\theta, \theta') = 2|\theta - \theta'| \).

**Definition 20.3 (Generalized Dudley Integral)** The generalized Dudley integral is

\[ J_q(D) = \int_0^D \chi_q^{-1}(\mathcal{N}(u, \mathbb{T}, \rho)) \, du \] (20.8)

where \( D = \text{supp}(\theta, \theta') \) is the diameter of \( \mathbb{T} \), \( \mathcal{N}(u, \mathbb{T}, \rho) \) is the \( \delta \)-covering number of \( \mathbb{T} \), and \( \chi_q^{-1}(y) = \left( \log(1 + y) \right)^{\frac{1}{q}} \).

Our main result for today is bounding the supremum of a \( \chi_q \) process by the generalized Dudley integral.

**Theorem 20.4** Let \( \{X_\theta, \theta \in \mathbb{T}\} \) be a \( \chi_q \) process with respect to \( \rho \). Then, \( \exists C > 0 \) such that

\[ \mathbb{P}\left( \sup_{\theta, \theta' \in \mathbb{T}} |X_\theta - X_{\theta'}| \geq C \left[ J_q(D) + \delta \right] \right) \leq 2 \exp\left\{ -\left( \frac{\delta}{D} \right)^q \right\} \] (20.9)

We will need the following lemma to prove this theorem.

**Lemma 20.5** Let \( Y_1, ..., Y_N \) be non-negative random variables s.t. \( \|Y\|_{\chi_q} \leq 1 \). Define, for a measurable set \( A \),

\[ \mathbb{E}_A(Y) := \int_A Y(\omega) \, dP(\omega) \quad \text{and} \]

\[ \mathbb{E}(Y|A) := \frac{\mathbb{E}_A(Y)}{P(A)} \] (20.10)
Then, for every measurable \( A \),

\[
\mathbb{E}_A(Y_i) \leq P(A)\chi_q^{-1}\left(\frac{1}{P(A)}\right) \quad \text{and} \quad (20.12)
\]

\[
\mathbb{E}_A(\max_{i=1,\ldots,N} Y_i) \leq P(A)\chi_q^{-1}\left(\frac{N}{P(A)}\right) \quad (20.13)
\]

**Proof:** (of Lemma) For the first statement, notice that

\[
\mathbb{E}_A(\chi_q(Y)) = \mathbb{E}_A(\chi_q(Y)||Y||_\chi) \leq \mathbb{E}_A(\chi_q(Y)) \leq ||Y||_\chi = 1. \quad \text{Therefore,}
\]

\[
\mathbb{E}_A(Y) = P(A)\mathbb{E}(Y|A) \quad \text{(20.14)}
\]

\[
= P(A)\chi_q^{-1}(\chi_q(Y)|A) \quad \text{since } Y \geq 0 \quad (20.15)
\]

\[
\leq P(A)\chi_q^{-1}\mathbb{E}(\chi_q(Y)|A) \quad \text{by the concavity of } \chi_q^{-1} \quad (20.16)
\]

\[
= P(A)\chi_q^{-1}\left(\frac{\mathbb{E}(\chi_q(Y))}{P(A)}\right) \quad (20.17)
\]

\[
\leq P(A)\chi_q^{-1}\left(\frac{1}{P(A)}\right) \quad \text{since } \mathbb{E}_A(\chi_q(Y)) \leq 1 \quad (20.18)
\]

For the second statement, begin by taking \( A_i = \{\omega : Y_i(\omega) = \max_{i=1,\ldots,N} Y_i\} \). Then,

\[
\int_A \max_{i=1,\ldots,N} Y_i(\omega)dP(\omega) = \sum_{i=1}^{N} \int_{A_i} Y_i(\omega)dP(\omega) \quad (20.20)
\]

\[
\leq \sum_{i=1}^{N} P(A_i)\chi_q^{-1}\left(\frac{1}{P(A_i)}\right) \quad (20.21)
\]

\[
= \sum_{i=1}^{N} P(A)\frac{P(A)}{P(A_i)}\chi_q^{-1}\left(\frac{1}{P(A_i)}\right) \quad (20.22)
\]

\[
\leq P(A)\chi_q^{-1}\left(\frac{N}{P(A)}\right) \quad \text{Jensen’s inequality for concave functions} \quad (20.23)
\]

With this lemma in hand, we can turn to proving our theorem.

**Proof:** (of Theorem) To begin with, we want to show that

\[
\mathbb{E}_A \left[|X_\theta - X_{\theta'}|_{\theta,\theta' \in \mathcal{T}}\right] \leq 8P(A)J_q(D) \quad (20.24)
\]

We will use a chaining argument very similar to the one used for Dudley’s (not generalized) method. Let \( U_m \) be a \( D2^{-m} \) minimal covering of \( \mathcal{T} \) such that \( |U_m| \leq N_m = N(D2^{-m}, \mathcal{T}, \rho) \). Let \( \pi_m : \mathcal{T} \to U_m \) be defined as \( \pi_m(\theta) = \arg\min_{\theta,\theta' \in U_m} \rho(\theta, \theta') \). Then,
\[
\mathbb{E}_A \left[ \sup_{\theta, \theta' \in T} |X_\theta - X_{\theta'}| \right] \leq 2 \sum_{m=1}^{\infty} \mathbb{E}_A \left[ \max_{\gamma \in U_m} |X_\gamma - X_{\pi_{m-1}(\gamma)}| \right] \tag{20.25}
\]

and for each \( \gamma \in U_m, \)

\[
\|X_\gamma - X_{\pi_{m-1}(\gamma)}\|_{\chi_q} \leq \rho(\gamma, \pi_{m-1}(\gamma)) \leq D^{-\left(m-1\right)} \tag{20.26}
\]

so by our lemma,

\[
\mathbb{E}_A \left[ \max_{\gamma \in U_m} |X_\gamma - X_{\pi_{m-1}(\gamma)}| \right] \leq P(A) D^{-\left(m-1\right)} \chi_q \left( \frac{N_m}{P(A)} \right) \tag{20.27}
\]

\[
\mathbb{E}_A \left[ \sup_{\theta, \theta' \in T} |X_\theta - X_{\theta'}| \right] \leq 2P(A) \sum_{m=1}^{\infty} D^{-\left(m-1\right)} \chi_q^{-1} \left( \frac{N_m}{P(A)} \right) \tag{20.28}
\]

\[
\leq cP(A) \int_0^D \chi_q^{-1} \left( \frac{N(u, T, \rho)}{P(A)} \right) du \tag{20.29}
\]

Now that we’ve bounded \( \mathbb{E}_A \left[ \sup_{\theta, \theta' \in T} |X_\theta - X_{\theta'}| \right], \) we need only to bound the (positive) deviation of \( \sup_{\theta, \theta' \in T} |X_\theta - X_{\theta'}| \) from its mean. We will need a slight variant of Markov’s inequality. Take some positive random variable \( Z, \) and let \( A \) be the event that \( Z > t. \) Then,

\[
P(A) = P(Z > t) \leq \frac{\mathbb{E}_A(Z)}{t} \tag{20.30}
\]

We also have that \( \chi_q^{-1}(st) \leq c \left[ \chi_q^{-1}(s) + \chi_q^{-1}(t) \right]. \) With these in mind, we proceed. From our previous work, we have that

\[
\mathbb{E}_A \left[ \sup_{\theta, \theta' \in T} |X_\theta - X_{\theta'}| \right] \leq 8J_q(D) \tag{20.31}
\]

Let \( Z = \sup_{\theta, \theta' \in T} |X_\theta - X_{\theta'}| \) and choose \( A = \{Z \geq t\}. \) Then

\[
P(A) \leq \frac{\mathbb{E}_A(Z)}{t} \tag{20.32}
\]

\[
\leq 8P(Z > t) \int_0^D \chi_q^{-1} \left( \frac{N(u, T, \rho)}{P(Z > t)} \right) du \rightarrow \tag{20.33}
\]

\[
t \leq 8c \left\{ J_q(D) + D\chi_q^{-1} \left( \frac{1}{P(Z > t)} \right) \right\} \tag{20.34}
\]
Finally, set $\delta > 0$ and let $t = 8c(J_q(D) + \delta)$, and we obtain,

$$
P(Z > 8c(J_q(D) + \delta)) \leq \frac{1}{\chi_q(\frac{\delta}{t})} 
$$

(20.35)

Getting from this to the final result will be a homework question.