**36-755: Advanced Statistical Theory** **Fall 2016**

# Lecture 20: November 7

*Lecturer: Alessandro Rinaldo* *Scribes: Alden Green*

**Note**: *LaTeX template courtesy of UC Berkeley EECS dept.*



## 20.1 Chaining and Orlicz Processes

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

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

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

Note that

$$\begin{aligned} \mathbb{P}(|X| > t) &= \mathbb{P}\left(\chi_q(\frac{|X|}{||X||_{\chi_q}}) > \chi_q \frac{t}{||X||\chi_q}\right) && \text{because } \chi_q \text{ increasing} \quad (20.2) \\ &\leq \frac{1}{\chi_q(\frac{t}{||X||_{\chi_q}})} && \text{by Markov} \quad (20.3) \end{aligned}$$

We will show on a later homework assignment that this implies

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

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:

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

**Remark 1** *If $\chi(u) = u^p, p \geq 1$, then*

$$||X||_{\chi} = (\mathbb{E}\left[|X|^p\right])^{\frac{1}{p}} \tag{20.6}$$

*More generally, any function $\chi : \mathbb{R}_+ \to \mathbb{R}_+$ strictly increasing, convex and with $\chi(0) = 0$ would yield a norm $||.||_{\chi}$ 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*

$$||X_\theta - X_{\theta'}||_{\chi_q} \leq \rho(\theta, \theta') \qquad \forall \theta, \theta' \in \mathbb{T} \tag{20.7}$$

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*

$$\mathcal{J}_q(D) = \int_0^D \chi_q^{-1}\left(\mathcal{N}(u, \mathbb{T}, \rho)\right) du \tag{20.8}$$

*where $D = \sup_{\theta,\theta'} \rho(\theta, \theta')$ is the diameter of $\mathbb{T}$, $\mathcal{N}(u, \mathbb{T}, \rho)$ is the u-covering number of $\mathbb{T}$, and $\chi_q^{-1}(y) = [\log(1 + y)]^{\frac{1}{q}}$.*

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*

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

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 $A$,*

$$\mathbb{E}_A(Y) := \int_A Y(\omega) dP(\omega) \qquad \textit{and} \tag{20.10}$$

$$\mathbb{E}(Y|A) := \frac{\mathbb{E}_A(Y)}{P(A)} \tag{20.11}$$

*Then, for every measurable* $A$*,*

$$\mathbb{E}_A(Y_i) \leq P(A)\chi_q^{-1}(\frac{1}{P(A)}) \qquad \textit{and} \tag{20.12}$$

$$\mathbb{E}_A(\max_{i=1,\ldots,N} Y_i) \leq P(A)\chi_q^{-1}(\frac{N}{P(A)}) \tag{20.13}$$

**Proof:** (of Lemma) For the first statement, notice that $\mathbb{E}_A\left(\chi_q(Y)\right) = \mathbb{E}_A\left(\chi_q(Y)\frac{||Y||_{\chi_q}}{||Y||_{\chi_q}}\right) \leq \mathbb{E}_A\left(\frac{\chi_q(Y)}{||Y||_{\chi_q}}\right) \leq ||Y||_{\chi_q} = 1$. Therefore,

$$\begin{aligned}
\mathbb{E}_A(Y) &= P(A)\mathbb{E}(Y|A) && (20.14)\\
&= P(A)\mathbb{E}(\chi_q^{-1}(\chi_q(Y))|A) \qquad \text{since } Y \geq 0 && (20.15)\\
&\leq P(A)\chi_q^{-1}\mathbb{E}(\chi_q(Y)|A) \qquad \text{by the concavity of } \chi_q^{-1} && (20.16)\\
&= P(A)\chi_q^{-1}(\frac{\mathbb{E}_A(\chi_q(Y))}{P(A)}) && (20.17)\\
&\leq P(A)\chi_q^{-1}(\frac{1}{P(A)}) \qquad \text{since } \mathbb{E}_A\left(\chi_q(Y)\right) \leq 1 && (20.18)\\
& && (20.19)
\end{aligned}$$

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

$$\begin{aligned}
\int_A \max_{i=1,\ldots,N} Y_i(\omega)dP(\omega) &= \sum_{i=1}^N \int_{A_i} Y_i(\omega)dP(\omega) && (20.20)\\
&\leq \sum_{i=1}^N P(A_i)\chi_q^{-1}(\frac{1}{P(A_i)}) && (20.21)\\
&= \sum_{i=1}^N P(A)\frac{P(A_i)}{P(A)}\chi_q^{-1}(\frac{1}{P(A_i)}) && (20.22)\\
&\leq P(A)\chi_q^{-1}(\frac{N}{P(A)}) \qquad \text{Jensen's inequality for concave functions} && (20.23)
\end{aligned}$$

■

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

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

$$\mathbb{E}_A\left[\underset{\theta,\theta'\in\mathbb{T}}{|X_\theta - X_{\theta'}|}\right] \leq 8P(A)\mathcal{J}_q(D) \tag{20.24}$$

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 $\mathbb{T}$ such that $|U_m| \leq N_m = \mathcal{N}(D2^{-m}, \mathbb{T}, \rho)$. Let $\pi_m : \mathbb{T} \to U_m$ be defined as $\pi_m(\theta) = \underset{\theta,\theta'\in U_m}{argmin}\rho(\theta,\theta')$. Then,

$$\mathbb{E}_A\left[\sup_{\theta,\theta'\in\mathbb{T}}|X_\theta - X_{\theta'}|\right] \le 2\sum_{m=1}^{\infty}\mathbb{E}_A\left[\max_{\gamma\in U_m}\left|X_\gamma - X_{\pi_{m-1}(\gamma)}\right|\right] \tag{20.25}$$

and for each $\gamma \in U_m$,

$$||X_\gamma - X_{\pi_{m-1}(\gamma)}||_{\chi_q} \le \rho(\gamma, \pi_{m-1}(\gamma)) \le D2^{-(m-1)} \tag{20.26}$$

so by our lemma,

$$\mathbb{E}_A\left[\max_{\gamma\in U_m}\left|X_\gamma - X_{\pi_{m-1}(\gamma)}\right|\right] \le P(A)D2^{-(m-1)}\chi_q\left(\frac{N_m}{P(A)}\right) \to \tag{20.27}$$

$$\mathbb{E}_A\left[\sup_{\theta,\theta'\in\mathbb{T}}|X_\theta - X_{\theta'}|\right] \le 2P(A)\sum_{m=1}^{\infty}D2^{-(m-1)}\chi_q^{-1}\left(\frac{N_m}{P(A)}\right) \tag{20.28}$$

$$\le cP(A)\int_0^D \chi_q^{-1}\left(\frac{\mathcal{N}(u,\mathbb{T},\rho)}{P(A)}\right)du \tag{20.29}$$

Now that we've bounded $\mathbb{E}_A\left[\sup_{\theta,\theta'\in\mathbb{T}}|X_\theta - X_{\theta'}|\right]$, we need only to bound the (positive) deviation of $\sup_{\theta,\theta'\in\mathbb{T}}|X_\theta - X_{\theta'}|$ 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,

$$\mathbb{P}(A) = \mathbb{P}(Z > t) \le \frac{\mathbb{E}_A(Z)}{t} \tag{20.30}$$

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

$$\mathbb{E}_A\left[\sup_{\theta,\theta'\in\mathbb{T}}|X_\theta - X_{\theta'}|\right] \le 8\mathcal{J}_q(D) \tag{20.31}$$

Let $Z = \sup_{\theta,\theta'\in\mathbb{T}}|X_\theta - X_{\theta'}|$ and choose $A = \{Z \ge t\}$. Then

$$P(A) \le \frac{\mathbb{E}_A(Z)}{t} \tag{20.32}$$

$$\le 8\frac{\mathbb{P}(Z>t)}{t}\int_0^D \chi_q^{-1}\left(\frac{\mathcal{N}(u,\mathbb{T},\rho)}{\mathbb{P}(Z>t)}\right)du \to \tag{20.33}$$

$$t \le 8c\left\{\mathcal{J}_q(D) + D\chi_q^{-1}\left(\frac{1}{\mathbb{P}(Z>t)}\right)\right\} \tag{20.34}$$

Finally, set $\delta > 0$ and let $t = 8c(\mathcal{J}_q(D) + \delta)$, and we obtain,

$$\mathbb{P}\left(Z > 8c(\mathcal{J}_q(D) + \delta)\right) \leq \frac{1}{\chi_q(\frac{\delta}{D})} \tag{20.35}$$

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