# Maximal Inequalities

Lab 6, BIO251

03/24/2014

## 1 Maximal Inequalities

Note that for $L_p$ norms we have the following:

$$\left|\left|\max_{1\le i\le m} X_i\right|\right|_p = \left(\mathrm{E}\max_{1\le i\le m}|X_i|^p\right)^{1/p} \le m^{1/p}\max_{1\le i\le m}||X_i||_p$$

In the last inequality we used the obvious bound: $\mathrm{E}\max_{1\le i\le m}|X_i|^p \le \sum_{i=1}^m \mathrm{E}|X_i|^p \le m\max_{1\le i\le m}\mathrm{E}|X_i|^p$.

We continue with a similar inequality for $\psi$ norms.

**Lemma 1** Let $\psi$ be a convex, nondecreasing, nonzero function with $\psi(0)=0$, and $\limsup_{x,y\to\infty}\frac{\psi(x)\psi(y)}{\psi(cxy)} < \infty$ for some constant $c$. Then for any random variables $X_1, X_2, \ldots, X_m$ we have:

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

where the constant $K$ depends on $\psi$ only.

**Proof.** First consider the case with $\psi$ satisfying the following constraint $\psi(x)\psi(y)\le\psi(cxy)$ for $x,y\ge 1$, and $\psi(1)\le\frac{1}{2}$. In this case we have $\psi(x/y)\le\psi(cx)/\psi(y)$ for $x\ge y\ge 1$. We then have for $y\ge 1$:

$$\begin{aligned}\max\psi\left(\frac{|X_i|}{Cy}\right) &\le \max\left[\frac{\psi(c|X_i|/C)}{\psi(y)} + \psi\left(\frac{|X_i|}{Cy}\right)\mathbb{1}\left(\frac{|X_i|}{Cy}<1\right)\right]\\ &\le \sum_i\frac{\psi(c|X_i|/C)}{\psi(y)} + \psi(1)\end{aligned}$$

Set $C = c\max||X_i||_\psi$, and take expectation:

$$\begin{aligned}\mathrm{E}\,\psi\left(\frac{\max|X_i|}{Cy}\right) &\le \sum_i\frac{\mathrm{E}\,\psi(|X_i|/\max||X_i||_\psi)}{\psi(y)} + \psi(1)\\ &\le \sum_i\frac{\mathrm{E}\,\psi(|X_i|/||X_i||_\psi)}{\psi(y)} + \psi(1)\\ &\le \frac{m}{\psi(y)} + \psi(1)\end{aligned}$$

For $y=\psi^{-1}(2m)$ we have that the above is $\le 1$. Note that $\psi^{-1}(2m)\ge\psi^{-1}(\frac{1}{2})\ge 1$. Therefore:

$$\left|\left|\max_{1\le i\le m} X_i\right|\right|_\psi \le \psi^{-1}(2m)c\max||X_i||_\psi$$
$$\le 2\psi^{-1}(m)c\max||X_i||_\psi$$

The last inequality follows from the following arguments: by the convexity of $\psi$ have $2m = \psi(\psi^{-1}(2m)) + \psi(0) \ge 2\psi(\frac{1}{2}\psi^{-1}(2m))$. Dividing by 2 and taking $\psi^{-1}$ gives $2\psi^{-1}(m) \ge \psi^{-1}(2m)$, which proves the inequality in this special case for $\psi$.

To finish the proof in the general case we need some more technical details, which are not essential. They are provided in the appendix but we don't discuss them in lab.

The important conclusion of the lemma is that for rapidly increasing $\psi$, e.g. $\psi(x) = \psi_p(x) = e^{x^p} - 1$ we have that the max norm is increasing the slowly because we have $\psi_p^{-1}(m) = (\log(1+m))^{1/p}$.

What if we have an infinite collection of random variables and we want to bound the supremum. Obviously the lemma above cannot handle this situation. Fortunately, Kolmogorov introduced a method with which people can deal with situations of that sort. The method is called *chaining*. The idea is that each of the variables in the supremum can be written as sum of "little links", and the bound depends on the size and the number of links needed. For a stochastic process $\{X_t : t \in T\}$ the number of links depends on the entropy of the indexing set (of the collection of random variables) for the semmimetric:

$$d(s,t) = ||X_s - X_t||_\psi$$

We next define what metric entropy would mean in our context.

**Definition (Covering numbers).** Let $(T,d)$ be a an arbitrary semimetric space. The *covering number* $N(\varepsilon, d)$ is the minimal number of balls of radius $\varepsilon$ needed to cover $T$.

**Definition (Packing numbers).** Call a collection of points $\varepsilon$-separated if the minimum distance between any two points is is strictly larger than $\varepsilon$. The *packing number* $D(\varepsilon, d)$ is the maximal number of $\varepsilon$-separated points in $T$.

There are corresponding *entropy numbers* which are the log's of the covering/packing numbers correspondingly. There is a close relationship between the packing and covering numbers and the following inequality makes it clear:

$$N(\varepsilon,d) \le D(\varepsilon,d) \le N(\frac{1}{2}\varepsilon, d)$$

To see the left hand side note that, if we position a ball of radius epsilon centered at each point of the maximal $\varepsilon$-separated collection of points, we get a covering of the set $T$. For the right hand side, observe that if we suppose the contrary, there are two points of the maximal $\varepsilon$-separated collection of points must lie in the same ball, which is a contradiction.

Any of the packing or covering numbers can be used for the present purposes. We will use the packing numbers $D$.

Obviously, both $N$ and $D$ increase when we decrease $\varepsilon$. By definition, the set $T$ is totally bounded, if both the covering and packing numbers are finite for every $\varepsilon > 0$. In the following theorem, the upper bound depends on the growth rate of $D$ in terms of $\varepsilon$ measured through an integral.

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

$$||X_s - X_t||_\psi \leq Cd(s,t), \text{ for every } s,t$$

for some semimetric $d$ on $T$ and a constant $C$. Then, for any $\eta, \delta > 0$,

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

for a constant $K$ depending on $\psi$ and $C$ only.

Here by separable processes, we mean a process such that $\sup_{d(s,t)<\delta} |X_s - X_t|$ remains almost surely the same if the index set $T$ is replaced by a suitable countable subset.

**Corollary.** The constant $K$ can be chosen such that:

$$\left|\left|\sup_{s,t} |X_s - X_t|\right|\right|_\psi \leq K \int_0^{\operatorname{diam} T} \psi^{-1}(D(\varepsilon,d))d\varepsilon$$

You can find the details how the corollary follows in the appendix. If we care about the sup of the process and not the increments we can still use the corollary above to show the following inequality for a fixed point $t_0$:

$$\left|\left|\sup_t |X_t|\right|\right|_\psi \leq ||X_{t_0}||_\psi + K \int_0^{\operatorname{diam} T} \psi^{-1}(D(\varepsilon,d))d\varepsilon$$

## 1.1 Sub-Gaussian Inequalities

As we mentioned last time, the standard normal distribution has tails satisfying the following inequality $P(|X| > x) \leq 2\exp(-x^2/2)$. In this section we discuss a class of random variables satisfying similar bounds.

Consider the following special case of Hoeffding's inequality:

**Lemma (Hoeffding's Inequality).** Let $a_1, \ldots, a_n$ are constants and $\varepsilon_1, \ldots, \varepsilon_n$ are iid*Rademacher* random variables: i.e. $P(\varepsilon = 1) = P(\varepsilon = -1) = 1/2$. Then we have:

$$P(|\sum \varepsilon_i a_i| > x) \leq 2e^{-\frac{1}{2}x^2/||a||^2}$$

for the Euclidean norm $||a||$. Consequently, $||\sum \varepsilon_i a_i||_{\psi_2} \leq \sqrt{6}||a||$.

**Proof.** We start by noting that the following:

$$\mathrm{E}\, e^{\lambda\varepsilon} = (e^\lambda + e^{-\lambda})/2 \leq e^{\lambda^2/2}$$

The last inequality can be seen upon a Taylor expansion: $(e^\lambda + e^{-\lambda})/2 = 1 + \lambda + \lambda^2/2 + \ldots + 1 - \lambda + \lambda^2/2 - \ldots = \sum_{i=0}^\infty \lambda^{2i}/(2i)! \leq \sum_{i=0}^\infty \lambda^{2i}/(2^i i!) = e^{\lambda^2/2}$. We then use Markov's inequality. For any $\lambda > 0$:

$$P(\sum \varepsilon_i a_i > x) = P(\exp(\lambda \sum \varepsilon_i a_i) > \exp(\lambda x)) \leq \frac{\mathrm{E}\exp(\lambda \sum \varepsilon_i a_i)}{\exp(\lambda x)}$$
$$\leq \exp(\lambda^2/2||a||^2 - \lambda x)$$

Minimizing for $\lambda$ gives $\lambda = -\frac{x}{||a||^2}$ to give a final bound of $e^{-\frac{1}{2}x^2/||a||^2}$. Similarly we take care of the other tail bound $P(\sum \varepsilon_i a_i < x)$, to give the bound form the Lemma statement.

Finally the $\psi_2$ norm bound follows from the last lemma last time, upon substituting $K = 1, C = \frac{1}{2}, p = 2$.

**Definition.** A stochastic process is called *sub-Gaussian* with respect to the semimetric $d$ on its index set if:

$$P(|X_s - X_t| > x) \leq 2\exp^{-\frac{1}{2}x^2/d^2(s,t)}, \text{ for every } s, t \in T, x > 0$$

Any Gaussian process is sub-Gaussian with respect to $d(s,t) = \sigma(X_s - X_t)$. Another example is the *Rademacher Process*:

$$X_a = \sum_{i=1}^n a_i \varepsilon_i, a \in \mathbb{R}^n$$

for Rademacher variables $\varepsilon_i, \dots, \varepsilon_n$. By Hoeffding's inequality, this is sub-Gaussian wrt to the Euclidean distance $d(a,b) = ||a - b||$.

Sub-Gaussian processes satisfy the increment bound $||X_s - X_t||_{\psi_2} \leq \sqrt{6}d(s,t)$. Since the inverse function of $\psi_2$ is (almost) the square root of the *log*, the entropy of the packing number appears in the integral. Consider the following corollary:

**Corollary.** Let $\{X_t : t \in T\}$ be a separable sub-Gaussian process. Then for any $\delta > 0$:

$$\mathrm{E} \sup_{d(s,t)<\delta} |X_s - X_t| \leq K \int_0^\delta \sqrt{\log(D(\varepsilon, d))} d\varepsilon$$

for a universal constant $K$. In particular, for any $t_0$:

$$\mathrm{E}\sup_t |X_t| \leq \mathrm{E}|X_{t_0}| + K \int_0^\infty \sqrt{\log(D(\varepsilon, d))} d\varepsilon$$

**Proof.** Apply the main theorem from today with $\psi_2(x) = \exp(x^2) - 1$, and $\eta = \delta$. We have $\psi_2^{-1}(x) = \sqrt{\log(1+x)}$. Note that $\psi_2^{-1}(x^2) \leq \sqrt{2}\psi_2^{-1}(x)$, for $x > 0$ because:

$$\sqrt{\log(1+x^2)} \leq \sqrt{\log(1+x)^2} = \sqrt{2}\sqrt{\log(1+x)}$$

Therefore the second part in the general Theorem – $\delta\psi^{-1}(D^2(\varepsilon, d)) \leq \sqrt{2}\delta\psi^{-1}(D(\varepsilon, d))$. Note that since this function is decreasing in $\eta$ it can be absorbed into the integral with the cost of increasing $K$ a bit. We then get:

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

Note that the maximum of $\delta$ that makes sense here is $\delta = \operatorname{diam} T$. For any $\varepsilon < \operatorname{diam} T$, we have that $D(\varepsilon, d) \geq 2$. Since $\log(1+m) \leq 2\log m$, we can discard the 1 inside the log by increasing $K$. Finally note that the $||\cdot||_{\psi_2} \geq ||\cdot||_2 \geq ||\cdot||_1$, as mentioned last time.

# A Some Details

**Lemma 1** We next show that there exist a constant $\sigma \leq 1$ and $\tau > 0$ such that $\phi(x) = \sigma\psi(\tau x)$ such that $\phi$ satisfies the two extra conditions that we required in the lemma, $\phi(1) \leq \frac{1}{2}, \phi(x)\phi(y) \leq \phi(cxy), x, y \geq 1$. First suppose that $\psi$ is bounded from above. Then select $\tau = 1$, $\sigma = \min(\frac{1}{2\psi(1)}, \frac{\psi(c)}{C^2})$, where $C$ is the upper bound for $\psi$. Direct verification shows that $\phi$ satisfies the restrictions.

Second, suppose that $\psi$ is unbounded. Consider the values $\sigma_n = \frac{1}{n}$ and $\tau_n = \psi^{-1}(\frac{n}{2})$. Direct verification shows that $\phi(1) = \frac{1}{2}$. Denote with $\phi_n(x) = \sigma_n\psi(\tau_n x)$ Suppose that for every $n$ we have that there exist $x_n \geq 1$ and $y_n \geq 1$ such that: $\phi_n(x_n)\phi_n(y_n) \geq \phi_n(cx_ny_n)$ or in other words we have $\frac{\psi(\tau_n x_n)\psi(\tau_n y_n)}{\psi(c\tau_n x_n \tau_n y_n)} \geq n$. Letting $n \to \infty$ would give a contradiction to the lim sup condition as $\tau_n x_n \geq \tau_n \to \infty$ and $\tau_n y_n \geq \tau_n \to \infty$.

Now apply the inequality for $\phi$:

$$\left|\left| \max_{1\leq i\leq m} X_i \right|\right|_\phi \leq 2\phi^{-1}(m)c \max ||X_i||_\phi$$

We need to translate this inequality in terms of $\psi$. We take advantage of the following inequality which we prove later:

$$||X||_\psi \leq ||X||_\phi/(\sigma\tau) \leq ||X||_\psi/\sigma \tag{1}$$

Using this we have:

$$\sigma\tau \left|\left| \max_{1\leq i\leq m} X_i \right|\right|_\psi \leq \left|\left| \max_{1\leq i\leq m} X_i \right|\right|_\phi \leq 2\phi^{-1}(m)c \max ||X_i||_\phi \leq \phi^{-1}(m)\tau c \max ||X_i||_\psi$$

The inequality for $\psi$ looks like:

$$\left|\left| \max_{1\leq i\leq m} X_i \right|\right|_\psi \leq \phi^{-1}(m)\frac{c}{\sigma} \max ||X_i||_\psi$$

Finally we show (1). First we show the left side inequality. We have, by convexity, that $\mathrm{E}\,\psi\left(\sigma\tau|X|/||X||_\phi\right) \leq \sigma\,\mathrm{E}\,\psi\left(\tau|X|/||X||_\phi\right) + \mathrm{E}(1-\sigma)\psi(0) = \mathrm{E}\,\phi(|X|/||X||_\phi) \leq 1$.

For the right side we have: $\mathrm{E}\,\phi(|X|/(\tau||X||_\psi)) = \mathrm{E}\,\sigma\psi(|X|/(||X||_\psi)) \leq \sigma \leq 1$

**Corrolary** Note here that when we take $\eta = \delta = \operatorname{diam} T$ we have that $D(\eta, d) = 1$. Therefore the second term in the Theorem $\delta\psi^{-1}(D^2(\eta, d)) = \delta\psi^{-1}(D(\eta, d))$. Since the function $\delta\psi^{-1}(D(\eta, d))$ is decreasing in $\eta$ we can absorb this term in the integral, by possibly increasing $K$ a bit.