Cours M2

Sub Gaussian random variables

If $X \sim N(0,1)$ and $p \in \mathbb{N}$

$$\mathbb{E} X^{2p} = (2p-1)(2p-3) \dots 1 \qquad \text{(integrate by parts)}$$

and therefore $$(\mathbb{E} X^{2p})^{\frac{1}{2p}} \underset{p \to \infty}{\sim} \sqrt{\frac{2p}{e}} \qquad \text{(STIRLING)}$$

If X is any r.v. one may define its subgaussian (or ψ_2) norm as

$$\|X\|_{\psi_2} = \sup_{p \geq 1} \frac{\|X\|_{L^p}}{\sqrt{p}}.$$

This is not the only definition ; one may define

$$\|X\|'_{\psi_2} = \inf \{ t > 0 \text{ s.t. } \mathbb{E} \exp((\tfrac{X}{t})^2) \leq 2 \}$$

and $\|\cdot\|_{\psi_2}$ and $\|\cdot\|'_{\psi_2}$ are equivalent norms.

$$\Rightarrow \quad \mathbb{P}(X \geq t \|X\|_{\psi_2}) \leq C \exp(-ct^2)$$

$\|\sum X_i\|_{\psi_2} \leq \sum \|X_i\|_{\psi_2}$

BETTER : if (X_i) are <u>independent</u> and <u>mean zero</u>

then $$\|\sum X_i\|_{\psi_2} \leq C \left(\sum \|X_i\|^2_{\psi_2}\right)^{\frac{1}{2}} \quad \text{for some } C > 0$$

implies Bernstein's inequalities.

Example
- Gaussian $\qquad X \sim N(0, \sigma^2) \Rightarrow \|X\|_{\psi_2} \leq C \|X\|_{L^2} = C\sigma$
- Bounded : ~~TF~~

If $(X_i)_{1 \leq i \leq N}$ ~~are~~ are subgaussian with $\|X_i\|_{\psi_2} \leq K$ then "union bound"

$$\mathbb{E} \sup X_i \leq CK\sqrt{\log N}$$

Recall that if $(X_t)_{t\in T}$ is a Gaussian process,
the index set T is naturally equipped with a distance

$$d(s,t) = \|X_t - X_s\|_{L^2}$$

This can be generalized to subgaussian processes.
Let (T,d) be a metric space. A collection of r.v. $(X_t)_{t\in T}$
is a subgaussian process with respect to d ~~with~~ with constant α if

$\mathbb{E}X_s = 0$ and $\forall s,t\in T \qquad \|X_s - X_t\|_{\psi_2} \leq \alpha\, d(s,t).$

Very often we want to estimate $\mathbb{E}\sup_{t\in T} X_t$ where X_t is a Gaussian or subGaussian process.

3 levels of sophistication

1. a ε-net argument + union bound
2. the chaining method (DUDLEY $\simeq$ all ε simultaneously)
3. generic chaining (TALAGRAND – much more complicated)

<u>Let's first see an example with 1</u>

A random vector $X : \Omega \to \mathbb{R}^n$ is <u>isotropic</u> if

$$\mathbb{E}X_i = 0 \quad \mathbb{E}X_iX_j = \delta_{i,j}$$

(the covariance matrix is identity)

<u>Fact</u> ~~Any~~ $\forall$ random vector X (not supported in an hyperplane) with $\mathbb{E}X=$
$\exists T$ ~~affine~~ linear s.t. $T(X)$ is isotropic.

A random vector $X : \Omega \to \mathbb{R}^n$ is subgaussian with constant C
if $\forall \theta\in S^{n-1}$

$$\|\langle X,\theta\rangle\|_{\psi_2} \leq C\, \underbrace{\|\langle X,\theta\rangle\|_{L^2}}_{=1 \text{ if } X \text{ is isotropic}}$$

<u>Example</u> : A random vector with independent subgaussian coordinates is subgaussian

<u>Proposition</u> Let $X_1, \dots X_m$ be independent subGaussian vectors in $\mathbb{R}^n$

$$A = \begin{bmatrix} X_1 & \cdots & X_m \end{bmatrix} \in \mathbb{R}^{n \times m}$$

$$\mathbb{P}(\|A\| \geq C \max(\sqrt{n}, \sqrt{m})) \leq \exp(-\min(n,m))$$

Consider $\mathcal{M} \subset S^{m-1}$, $\mathcal{N} \subset S^{n-1}$ ε-nets for the Euclidean distance

(can have $|\mathcal{M}| \leq \left(\frac{3}{\varepsilon}\right)^m$, $|\mathcal{N}| \leq \left(\frac{3}{\varepsilon}\right)^n$)

$$\|A\| = \sup_{\substack{x \in S^{m-1} \\ y \in S^{n-1}}} \langle Ax, y \rangle$$

For $x \in S^{m-1}$, $y \in S^{n-1}$ let $x_0 \in \mathcal{N}$, $y_0 \in \mathcal{M}$ s.t. $\|x - x_0\| \leq \varepsilon$, $|y - y_0| \leq \varepsilon$

$$\langle Ax, y \rangle = \underbrace{\langle Ax_0, y_0 \rangle}_{\leq \sup_{x_0 \in}} + \underbrace{\langle Ax_0, y - y_0 \rangle}_{\leq \varepsilon \|A\|} + \underbrace{\langle A(x - x_0), y \rangle}_{\leq \varepsilon \|A\|}$$

take $\sup_{x,y}$ $\quad \|A\| \leq \sup_{\substack{x_0 \in \mathcal{M} \\ y_0 \in \mathcal{N}}} \langle Ax_0, y_0 \rangle + 2\varepsilon \|A\|$

$$\|A\| \leq \frac{1}{1 - 2\varepsilon} \sup_{\substack{x_0 \in \mathcal{M} \\ y_0 \in \mathcal{N}}} \langle Ax_0, y_0 \rangle \qquad \text{choose } \varepsilon = \frac{1}{4}$$

$$\mathbb{P}(\|A\| \geq t) \leq 12^{m+n} \, \mathbb{P}\left(\langle Ax_0, y_0 \rangle \geq \frac{t}{2}\right)$$

$$\leq 12^{m+n} \, C \exp(-ct^2)$$ and choose $t \sim \max(\sqrt{m}, \sqrt{n})$

~~Remarks~~ In the same way one can prove that subgaussian matrices satisfy the RIP property

The SUDAKOV minoration

If (X_t) is a Gaussian process then

$$\sup_{\varepsilon>0} \varepsilon \sqrt{\log N(T,d,\varepsilon)} \leq C\, \mathbb{E} \sup X_t$$

Proof $N(T,d,\varepsilon) \leq P(T,d,\frac{\varepsilon}{2})$

Let $t_1 \dots t_N$ be $\frac{\varepsilon}{2}$-separated with $N = N(T,d,\varepsilon)$

$$d(t_i,t_j) = \| X_{t_i} - X_{t_j}\|_{L^2} \geq \varepsilon = \| \tfrac{\varepsilon}{\sqrt{2}} Y_i - \tfrac{\varepsilon}{\sqrt{2}} Y_j \|_{L^2}$$

Let Y_i be iid $N(0,1)$ variables

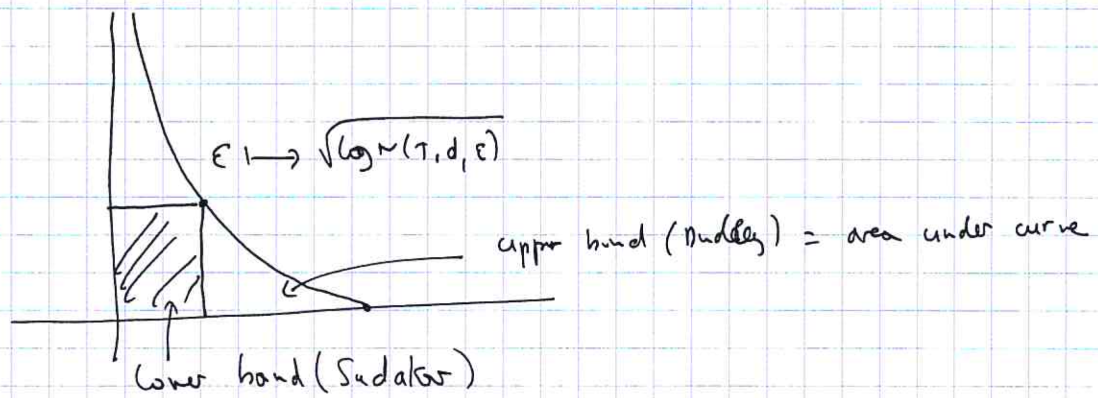
~~On peut donc~~
We can apply SLEPIAN's Lemma to the process $(\frac{\varepsilon}{\sqrt{2}} Y_i)$ and (X_{t_i})

$$\frac{\varepsilon}{\sqrt{2}}\sqrt{2\log N} \leq \mathbb{E}\sup_i Y_{t_i} \leq \mathbb{E}\sup_i X_{t_i} \leq \mathbb{E}\sup_{t\in T} X_t . \quad \blacksquare$$

There is an upper bound in similar spirit: the DUDLEY majoration

Theorem (DUDLEY's inequality) Let (T,d) be a compact metric space. If $(X_t)_{t\in T}$ is a subgaussian process with constant α with respect to d then

$$\mathbb{E}\sup_{t\in T} X_t \leq C\alpha \int_0^{\operatorname{diam}(T)\ (\infty)} \sqrt{\log N(T,d,\varepsilon)}\, d\varepsilon$$



Proof of DUDLEY inequality: the chaining method

We may assume $\alpha = 1$. Subgaussianity implies

$$\mathbb{P}(X_s - X_t > x) \le C \exp\left(- c \frac{x^2}{d(s,t)^2} \right)$$

We show the equivalent bound

$$\mathbb{E} \sup_t X_t \le C \sum_{k \in \mathbb{Z}} 2^{-k} \sqrt{\log N(T, d, 2^{-k})} \qquad \text{(compare series to integral)}$$

We can assume T finite

~~$\forall t \in T$~~ $\forall k \in \mathbb{Z}$ Let $\mathcal{N}_k$ be a ε_k-net in (T,d) with $\mathrm{card}(\mathcal{N}_k) = N(T, d, \varepsilon_k)$

$\varepsilon_k = 2^{-k}$

k_{max} = minimal k s.t. $\mathcal{N}_k = T$

k_{min} = maximal k s.t. $\mathrm{card}(\mathcal{N}_k) = 1$. $\quad \mathcal{N}_{k_{min}} = \{t_0\}$.

$\forall t \in T$ $\forall k \in \mathbb{Z}$ Choose $\pi_k(t)$ such that $d(t, \pi_k(t)) \le \varepsilon_k$.

$$\mathbb{E} \sup_t X_t = \mathbb{E} \sup_t (X_t - X_{t_0})$$

<u>chaining equation</u>

$$X_t - X_{t_0} = X_{\pi_{k_{max}}(t)} - X_{\pi_{k_{min}}(t)} = \sum_{k=k_{min}}^{k_{max}-1} X_{\pi_{k+1}(t)} - X_{\pi_k(t)}$$

so

$$\mathbb{E} \sup_t (X_t - X_{t_0}) \le \sum_{k=k_{min}}^{k_{max}-1} \underbrace{\mathbb{E} \sup_{t \in T} \left[X_{\pi_{k+1}(t)} - X_{\pi_k(t)} \right]}$$

this is the supremum of $\le \mathrm{card}(\mathcal{N}_k)\, \mathrm{card}(\mathcal{N}_{k+1})$ subgaussian r.v. with constant

$$d(\pi_{k+1}(t), \pi_k(t)) \le \varepsilon_k + \varepsilon_{k+1} \le 2 \cdot 2^{-k}$$

So $\mathbb{E} \sup_{t \in T} \left[X_{\pi_{k+1}(t)} - X_{\pi_k(t)} \right] \leq C \epsilon_k \sqrt{\log[\mathrm{card}(N_k)\,\mathrm{card}(N_{k+1})]}$

$\leq C\, \epsilon_{k+1} \sqrt{\log \mathrm{card}\, N_{k+1}}$

as needed. ∎

The same proof gives a "with high probability" estimation.

Application : concentration of subspace.

Let ~~a~~ $0 < k < n$ be integers

$G_{n,k} = \{ k\text{-dimensional subspaces of } \mathbb{R}^n \}$. (grassmann manifold)

There is a natural probability measure on $G_{n,k}$, which is the distribution of $\mathrm{span}(X_1, \dots X_k)$ for $X_1 \dots X_k$ iid $N(0, \mathrm{Id}_n)$

Let $\mu_{n,k}$ this measure

~~Proposition : Let $f : S^{n-1} \to \mathbb{R}$ be a 1-Lipschitz function and $E \in G_{n,k}$ chosen at random with distribution $\mu_{n,k}$.~~

Here is another approach.

~~the~~ The orthogonal group $O(n)$ has a unique probability measure μ ~~such that~~ which is invariant under right and left multiplication: the Haar measure (If $O \sim \mu$ then $\forall A \in O(n)$ $OA \sim \mu$, $AO \sim \mu$).

$\forall E \in G_{n,k}$

If $O \in O(n)$ is Haar distributed then $O(E) \sim \mu_{n,k}$

($\mu_{n,k}$ is the unique probability measure on $G_{n,k}$ invariant under the action of $O(n)$)

Proposition Let $f : S^{n-1} \to \mathbb{R}$ be a 1-Lipschitz function, $\varepsilon > 0$ and ~~$O \in O(n)$ Haar-distributed~~ $E \in G_{n,k}$ with distribution $\mu_{n,k}$. Then with high probability

if $k \leq c n \varepsilon^2$ then $\sup_{x \in E \cap S^{n-1}} |f(x) - \mathbb{E} f| \leq \varepsilon$

~~A~~ Proof Assume $\mathbb{E} f = 0$ and consider

$E = O(E_0)$ for some fixed $E_0 \simeq \mathbb{R}^k$

Consider the process $X_t = f(O(t))$ indexed by $t \in S^{k-1}$.

Claim $\mathbb{P}(X_s - X_t > \lambda) \leq \exp\left(-\frac{(n-1)\lambda^2}{2|s-t|^2}\right)$

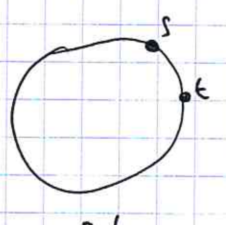
so the process $(X_t)_{t \in S^{k-1}}$ is subgaussian with constant $\frac{C}{\sqrt{n}}$

Dudley $\Rightarrow$ $\mathbb{E} \sup_{t \in S^{k-1}} X_t \leq \frac{C}{\sqrt{n}} \int_0^{1/2} \underbrace{\sqrt{\log N(S^{k-1}, |\cdot|, \delta)}}_{\leq \left(\frac{C}{\delta}\right)^k} d\delta$

$$\leq \frac{C}{\sqrt{n}} \int_0^{1/2} \sqrt{k} \sqrt{\log \frac{1}{\delta}} \, d\delta \leq \underbrace{C\sqrt{\frac{k}{n}}}_{\leq \varepsilon}.$$

for $k \leq C n \varepsilon^2$

Proof of the claim



$x = \frac{s+t}{2}$ $y = \frac{s-t}{2}$ ~~$x = \frac{s+t}{2}$~~ $s = x+y$, $t = x-y$

Conditioned on $\underset{u}{O(x)}$, $\overset{v}{O\left(\frac{y}{|y|}\right)}$ is distributed according to the measure σ on $S^{n-2} \subset u^{\perp}$

$O(s) = u + |y| v$
$O(t) = u - |y| v$

$v \mapsto \ldots$ is a $2|y|$-Lipschitz function

$$\mathbb{P}(X_s - X_t > \lambda) = \mathbb{P}(f(O(s)) - f(O(t)) > \lambda) = \mathbb{P}(f(u+|y|v) - f(u-|y|v) > \lambda)$$