On weakly bounded empirical processes

Shahar Mendelson

Technion and ANU

THE AUSTRALIAN NATIONAL UNIVERSITY

Questions

Let F be a set of mean-zero functions on (Ω, μ) and let $\sigma = (X_1, ..., X_N)$ be independent, distributed according to μ.

Set

$$P_\sigma F = \left\{ \, (f(X_i))_{i=1}^N : \; f \in F \, \right\}$$

the coordinate projection of F onto σ.

Questions II

What is the structure of a typical

$$P_\sigma F = \left\{ \, (f(X_i))_{i=1}^N : \; f \in F \, \right\}?$$

If $\phi : \mathbb{R} \to \mathbb{R}$ is a reasonable function, is

$$\sup_{f \in F} \left| \frac{1}{N} \sum_{i=1}^N \phi(f)(X_i) - \mathbb{E}\phi(f) \right|$$

small, and if so, why?

- $\phi(t) = t \implies$ Uniform law of large numbers
- $\phi(t) = t^2 \implies$ Uniform CLT

Example I

Let $F = \{\langle x, \cdot \rangle : x \in S^{n-1}\}$, μ an isotropic measure on $\mathbb{R}^n$. Set

$$\Gamma = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \langle X_i, \cdot \rangle e_i.$$

Then,

$$P_\sigma F = \left\{ \left(\langle X_i, x \rangle \right)_{i=1}^{N} : x \in S^{n-1} \right\} = \sqrt{N}\Gamma(S^{n-1}),$$

and

$$\sup_{f \in F} \left| \frac{1}{N} \sum_{i=1}^{N} f^2(X_i) - \mathbb{E} f^2 \right| = \sup_{x \in S^{n-1}} \left| |\Gamma x|^2 - 1 \right| = (*).$$

So why is (*) small?

Example II

If μ is L-subgaussian, i.e.,

$$\|\langle X, x\rangle\|_{\psi_2} \leq L|x|$$

then if $x \in S^{n-1}$, with probability $1 - 2\exp(-c_1\varepsilon^2 N)$,

$$\left|\frac{1}{N}\sum_{i=1}^{N}\langle X_i, x\rangle^2 - 1\right| < \varepsilon.$$

On the other hand, the "complexity" of S^{n-1} is $\exp(c_2 n)$.

A good estimate – when $N \geq c_3(\varepsilon)n$, and the rate of convergence is

$$\sim \sqrt{\frac{n}{N}} + \frac{n}{N}.$$

Example III

What happens when μ is isotropic, log-concave? Then

- $\|\langle X, x \rangle\|_{\psi_1} \leq L|x|$, but $\|\langle X, x \rangle\|_{\psi_2}$ can be very large ($\sim \sqrt{n}|x|$).

- Concentration: since $\langle X, x \rangle^2 \in L_{\psi_{1/2}}$, the degree of concentration of empirical means is $\exp(-c_1(\varepsilon)\sqrt{N})$.

- Complexity: $\exp(c_2 n)$.

One can expect a good estimate when $N \geq c_3(\varepsilon) n^2$.

Proved by Kannan–Lovász–Simonovits (97) under more general assumptions.

Example IV

Bourgain (98): it is enough to take $N \sim c_1(\varepsilon) n \log^3 n$.

Note:

$\sup_{x \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^N \langle X_i, x \rangle^2 - 1 \right|$ is small NOT because of individual concentration.

Partial history of the progress:

Rudelson (99), Giannopoulos–Milman (00), Giannopoulos–Hartzoulaki–Tsolomitis (05), Paouris (06), Guédon–Rudelson (07), Aubrun (07), M (08), Adamczak–Litvak–Pajor–Tomczak-Jaegermann (10),

showing that $N \sim c_2(\varepsilon) n$ is enough to ensure that

$$\sup_{x \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^N \langle X_i, x \rangle^2 - 1 \right| < \varepsilon.$$

Observations

The proofs are restricted:

- Use linear structure of the problem.
- Use that the indexing set is entire sphere – based either on uniform estimates on $\left|\sum_{i \in I} X_i\right|$ or on a noncommutative Khintchine inequality.
- If $T \subset S^{n-1}$ there is no corresponding estimate for $\sup_{t \in T} \left| \frac{1}{N} \sum_{i=1}^{N} \left\langle X_i, t \right\rangle^2 - 1 \right|$.
- Even the [ALPT] estimate is off by $\log N$ for $N \geq c(\beta) n^{1+\beta}$ for any $\beta > 0$ (rate estimate of $\sim \sqrt{\frac{n}{N}} \log(eN/n)$).

What happens for a general class?

General results

The complexity parameter I will use is

$$\gamma_2(F,\psi_2)=\inf\sup_{f\in F}\sum_{s=0}^{\infty}2^{s/2}d_{\psi_2}(f,\pi_s(f)). \qquad (*)$$

If the $L_2(\mu)$ and the $\psi_2(\mu)$ metrics are equivalent (F is subgaussian) then

$$\gamma_2(F,\psi_2)\sim\mathbb{E}\sup_{f\in F}G_f$$

.

A simple chaining argument shows that

$$\sup_{f\in F}\left|\frac{1}{N}\sum_{i=1}^{N}f(X_i)-\mathbb{E}f\right|\lesssim\frac{\gamma_2(F,\psi_2)}{\sqrt{N}}.$$

And a similar estimate is not true for any ψ_α metric, for $\alpha<2$!

$\sup_{f \in F} \left| \frac{1}{N} \sum_{i=1}^{N} f^2(X_i) - \mathbb{E} f^2 \right|$??

Contraction (simple):

$$\sup_{f \in F} \left| \frac{1}{N} \sum_{i=1}^{N} f^2(X_i) - \mathbb{E} f^2 \right| \lesssim \sup_{f \in F} \|f\|_\infty \cdot \frac{\gamma_2(F, \psi_2)}{\sqrt{N}}.$$

M–Pajor–Tomczak-Jaegermann (07):

$$\sup_{f \in F} \left| \frac{1}{N} \sum_{i=1}^{N} f^2(X_i) - \mathbb{E} f^2 \right| \lesssim \sup_{f \in F} \|f\|_{\psi_2} \cdot \frac{\gamma_2(F, \psi_2)}{\sqrt{N}} + \frac{\gamma_2^2(F, \psi_2)}{N}.$$

This is good enough for many geometric applications (e.g., low-M^* estimates for subgaussian operators, reconstruction using subgaussian measurements, norms of subgaussian operators into ℓ_2^N....), but it is NOT good enough to handle log-concave operators.

The main result

If F is class of zero-mean functions then

$$\mathbb{E}\sup_{f\in F}\left|\frac{1}{N}\sum_{i=1}^{N} f^2(X_i)-\mathbb{E}f^2\right| \lesssim \sup_{f\in F}\|f\|_{\psi_1}\cdot\frac{\gamma_2(F,\psi_2)}{\sqrt{N}}+\frac{\gamma_2^2(F,\psi_2)}{N}.$$

and a similar bound holds with high probability (though with a weaker probability estimate than in the subgaussian case).

Main ideas of the proof

Step I: structural results on F:

with probability at least $1 - 2\exp(-c_1 t \log N)$, for every $I \subset \{1, ..., N\}$ and every $f \in F$,

$$\left(\sum_{i \in I} f^2(X_i) \right)^{1/2} \lesssim_t \gamma_2(F, \psi_2) + d_{\psi_\alpha} \sqrt{|I|} \log^{1/\alpha} (eN/|I|) \, ,$$

where $d_{\psi_\alpha} = \sup_{f \in F} \|f\|_{\psi_\alpha}$, and this estimate is optimal.

In particular,

$$P_\sigma F \subset ct \left(\gamma_2(F, \psi_2) B_2^N + d_{\psi_\alpha} B_{\psi_\alpha^N} \right) .$$

Step I

To put Step I in context, note that if $Y \in L_{\psi_\alpha}$ then with high probability, a vector of independent copies of Y satisfies

$$(Y_1, ..., Y_N) \in \|Y\|_{\psi_\alpha} B_{\psi_\alpha^N} \Longleftrightarrow Y_i^* \lesssim \|Y\|_{\psi_\alpha} \log^{1/\alpha}(eN/i).$$

In other words, $P_\sigma F$ has

1. A "peaky" part – originates from the complexity of F, has a short support and is nicely bounded in ℓ_2^N.
2. A "regular" part – as if F has a ψ_α envelope function.

So, where do we stand?

This gives us a picture that is almost optimal:

$$\sup_{f \in F} \frac{1}{N} \sum_{i=1}^{N} f^2 \mathbb{1}_{\{|f| \geq \square\}}(X_i) \lesssim \frac{\gamma_2^2(F, \psi_2)}{N},$$

and

$$\sup_{f \in F} \left| \frac{1}{N} \sum_{i=1}^{N} f^2 \mathbb{1}_{\{|f| < \square\}}(X_i) - \mathbb{E} f^2 \mathbb{1}_{\{|f| < \square\}} \right| \lesssim \square \sup_{f \in F} \left| \frac{1}{N} \sum_{i=1}^{N} f(X_i) - \mathbb{E} f \right| .$$

So, the process is small because:

• The L_2^N norms of the "peaky" parts of functions in F are uniformly small – BUT THERE IS NO CONCENTRATION!!!

• And for the bounded part, empirical means do concentrate around true means at the correct rate.

So, where do we stand II?

This is NOT enough to prove the result!

$$\sup_{f \in F} \frac{1}{N} \sum_{i=1}^{N} f^2 \mathbb{1}_{\{|f| \geq \square\}}(X_i) \lesssim \frac{\gamma_2^2(F, \psi_2)}{N},$$

and

$$\sup_{f \in F} \left| \frac{1}{N} \sum_{i=1}^{N} f^2 \mathbb{1}_{\{|f| < \square\}}(X_i) - \mathbb{E} f^2 \mathbb{1}_{\{|f| < \square\}} \right| \lesssim \square \sup_{f \in F} \left| \frac{1}{N} \sum_{i=1}^{N} f(X_i) - \mathbb{E} f \right|.$$

However, $\square \neq d_{\psi_1}$, but rather

$$\square \sim d_{\psi_1} \log^{1/\alpha} \left(N d_{\psi_1}^2 / \gamma_2^2(F, \psi_2) + 2 \right).$$

off by a $\log N$ factor for "large" N - where one expects to see concentration.

Truncation methods cannot overcome this!!

Step II - Highlights

Consider the Bernoulli process

$$\sup_{f \in F} \sum_{i=1}^N \varepsilon_i f^2(X_i) = \sup_{v \in P_\sigma F} \sum_{i=1}^N \varepsilon_i v_i^2 .$$

To preform chaining, at the s-th step one needs to control $2^{2^{s+1}}$ points.

Observe: with probability at least $1 - 2\exp(-t^2/2)$,

$$\left| \sum_{i=1}^N \varepsilon_i a_i \right| \leq \sum_{i=1}^{\ell} a_i^* + \sqrt{t} \left(\sum_{i=\ell+1}^N (a^2)_i^* \right)^{1/2} .$$

In our case, $v = (f(X_i))_{i=1}^N$, $u = (g(X_i))_{i=1}^N$ and

$$a_i = (v_i^2 - u_i^2) = (v_i - u_i)(v_i + u_i).$$

Step II - Highlights

• Use the "global information" on a monotone rearrangement of the coordinates of $(u+v)_{i=1}^N$ from the previous step: $u,v \in P_\sigma F$ implying decomposition to "short support" + bounded in ℓ_2^N, and "regular part".

$$P_\sigma F \subset c(t)\left(\gamma_2(F,\psi_2)B_2^N + d_{\psi_1}B_{\psi_1^N}\right).$$

• Obtain a similar – but "local" information on a monotone rearrangement of $2^{2^{s+1}}$ vectors $(v_i-u_i)_{i=1}^N = ((f-g)(X_i))_{i=1}^N$ – that depends on $\|f-g\|_{\psi_2}$, $\|f-g\|_{\psi_1}$ and on s.

• For each vector, select ℓ in $\sum_{i=1}^{\ell} a_i^* + \sqrt{t}\left(\sum_{i=\ell+1}^{N}(a^2)_i^*\right)^{1/2}$ according to the above information....

• Now, completing the chaining argument concludes the proof.

Outcomes I

• Many classical results from AGA hold for a random log-concave operator, with the complexity parameter being $\gamma_2(F, \psi_2)$ rather than the Gaussian parameter $\sqrt{n}M^*$. For example, low-M^* estimates, bounds on the norm $\|\Gamma\|_{K \to \ell_p^N}$, etc.

• Question: How does one estimate $\gamma_2(S^{n-1}, \psi_2(\mu))$ for an isotropic, log-concave measure μ?

• If $\bar{\mu}$ is the conditioning of μ to a "large subset" of $\mathbb{R}^n$, (e.g., to $c\sqrt{n}B_2^n$), then

$$\gamma_2(S^{n-1}, \psi_2(\bar{\mu})) \lesssim \sqrt{n \log n}.$$

If μ is unconditional, then

$$\gamma_2(S^{n-1}, \psi_2(\bar{\mu})) \lesssim \sqrt{n}.$$

Outcomes II

This is enough to show that if $N \leq \exp(\sqrt{n})$, then with high μ^N-probability,

$$\sup_{x \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \langle X_i, x \rangle^2 - 1 \right| \lesssim \sqrt{\frac{n \log n}{N}} + \frac{n \log n}{N},$$

and if μ is unconditional and $N \leq n^{\alpha}$, then with high μ^N-probability

$$\sup_{x \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \langle X_i, x \rangle^2 - 1 \right| \lesssim_{\alpha} \sqrt{\frac{n}{N}} + \frac{n}{N}.$$

We should expect the subgaussian rate, at least when $N \leq \exp(\sqrt{n})$.

Outcomes III

The fact that

$$\mathbb{E}\sup_{f\in F}\left|\frac{1}{N}\sum_{i=1}^{N} f^2(X_i)-\mathbb{E}f^2\right| \lesssim \sup_{f\in F}\|f\|_{\psi_1}\cdot\frac{\gamma_2(F,\psi_2)}{\sqrt{N}}+\frac{\gamma_2^2(F,\psi_2)}{N}.$$

is another good reason to study $\gamma_2(F,\psi_2)$ and extend the beautiful theory of $\gamma_2(F,L_2)$.

But, in fact, the complexity parameter we need is really better:

$$\inf\sup_{f\in F}\sum_{s=0}^{\infty}\|f-\pi_s f\|_{L_{2^s}} \leq \inf\sup_{f\in F}\sum_{s=0}^{\infty} 2^{s/2}\|f-\pi_s f\|_{\psi_2} (!!)$$