$\boxed{C}$ Logarithmic-Sobolev inequalities (contd.)

(b) Gaussian log-Sobolev inequality

Theorem (Gross '75)

Let $X_1, \dots, X_n$ be iid $\mathcal{N}(0,1)$. For a bounded, continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, we have

$$(\#) \qquad \mathrm{Ent}(f^2) \le 2\,\mathbb{E}\left[\|\nabla f\|_2^2\right].$$

(This looks very similar to the Gaussian Poincaré inequality, but is stronger than that (why?))

Proof. The proof is very similar to that of Gaussian Poincare.
First, we note that it suffices to show (#) for $n=1$. Indeed, using the sub-additivity of Ent,

$$\mathrm{Ent}(f^2) \le \sum_{i=1}^n \mathbb{E}\left[\mathrm{Ent}_{(i)}(f^2)\right]$$

(by (#) for $n=1$)
$$\le \sum_{i=1}^n \mathbb{E}\left[2\mathbb{E}_{(i)}\left|\frac{\partial f}{\partial x_i}(\tilde{X}^{(i)})\right|^2\right]$$
$$= 2\,\mathbb{E}\left[\|\nabla f(X)\|_2^2\right]$$

Next, focusing on $n=1$, let $\varepsilon_1, \dots, \varepsilon_m$ be iid unif$\{-1,+1\}$.
Let $g(\varepsilon_1, \dots, \varepsilon_m) \stackrel{\text{def}}{=} f\left(\frac{1}{\sqrt{m}} \sum_{i=1}^m \varepsilon_i\right)$.

By the binary log-Sobolev inequality, ②

$$\mathrm{Ent}(g^2) \leq 2\mathcal{E}(g),$$

where

$$\mathrm{Ent}(g^2) = \mathbb{E}\left[f^2\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\xi_i\right)\log f^2\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\xi_i\right)\right] - \mathbb{E}\left[f^2\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\xi_i\right)\right]\log \mathbb{E}\left[f^2\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\xi_i\right)\right]$$

$$\mathcal{E}(g) = \frac{1}{4}\sum_{i=1}^{m} \mathbb{E}\left(\left(f\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\xi_j\right) - \underbrace{f\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\xi_j - \frac{2}{\sqrt{m}}\xi_i\right)}_{g(\bar{\xi}^{(i)})}\right)^2\right]$$

Note that

$$f\left(\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\xi_j\right) - f\left(\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\xi_j - \frac{2\xi_i}{\sqrt{m}}\right)$$

$$\leq \frac{2\xi_i}{\sqrt{m}} \max_{\theta\in\left(\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\xi_j - \frac{2\xi_i}{\sqrt{m}},\ \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\xi_i\right)} |f'(\theta)|$$

Therefore,

$$\overset{\text{def}}{=} \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\xi_i$$

$$\mathrm{Ent}(g^2) \leq \frac{2}{m}\sum_{i=1}^{m}\mathbb{E}\left[\max_{\theta\in\left(X_m - \frac{2}{\sqrt{m}},\ X_m + \frac{2}{\sqrt{m}}\right)} |f'(\theta)|^2\right]$$

Taking limit $m \to \infty$, the central limit theorem yields

$$\mathrm{Ent}(f^2) \leq 2\,\mathbb{E}\left[|f'(X)|^2\right]$$

where

$$X_m = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\xi_i \xrightarrow{d} X \sim \mathcal{N}(0,1),$$

which completes the proof. ∎

<u>Corollary</u> (Gaussian Concentration: Tsirelson–Ibragimov–Sudakov ineq.)

For a Lipchitz function $f: \mathbb{R}^n \to \mathbb{R}$ and $(X_1, \ldots, X_n)$ iid $N(0,1)$,

$$\mathbb{P}\left(f(X) \geq \mathbb{E}[f(X)] + t\right) \leq e^{-t^2/2}.$$

(f is Lipchitz if $|f(x)-f(y)| \leq \|x-y\|$ $\forall\, x,y$. Suppose f is L-Lipchitz, i.e., $|f(x)-f(y)| \leq L\|x-y\|$ $\forall\, x,y$. Then, replacing f with $L^{-1} \cdot f$ in the ineq. above, we get

$$\mathbb{P}\left(f(X) \geq \mathbb{E}[f(X)] + t\right) \leq e^{-t^2/2L^2}.)$$

<u>Proof.</u> First, assume f is differentiable with $\|\nabla f(x)\| \leq 1$. Then,

$$\begin{aligned}
\mathrm{Ent}(e^{\lambda f}) &\leq 2\,\mathbb{E}\left[\|\nabla e^{\frac{\lambda}{2} f(X)}\|_2^2\right] \quad \text{(by GLS)} \\
&= \frac{\lambda^2}{2}\,\mathbb{E}\left[e^{\lambda f(X)} \cdot \|\nabla f(X)\|_2^2\right] \\
&\leq \frac{\lambda^2}{2}\,\mathbb{E}[e^{\lambda f(X)}]
\end{aligned}$$

Thus, the claim follows by Herbst's argument.

In general, we approximate f with a smooth function. Specifically, let $g_m(x) = \sqrt{\frac{m}{2\pi}}\, e^{-\frac{m\|x\|_2^2}{2}}$ (Gaussian Kernel)

Consider $f_m(x) = (f * g_m)(x)$. Then,

$$\lim_{m \to \infty} f_m(x) = f(x) \quad \text{and} \quad |f_m(x) - f_m(y)| \leq \|x-y\| \int g_m(z)\,dz.$$

Thus, $f_m(x)$ is Lipchitz for all m, and the sequence $\{f_m\}_{m=1}^{\infty}$ converges pointwise to f. Further, f_m is differentiable. Therefore, $\left| f_m(x) - f_m(x+tv) \right| \leq t \quad \forall\, v$ s.t. $\|v\| = 1$.

$$\Rightarrow \quad \left| \nabla f_m(x) \cdot v \right| \leq 1 \quad \forall\, v \text{ s.t. } \|v\| = 1$$

$$\Rightarrow \quad \|\nabla f_m(x)\| \leq 1 \quad \forall\, x.$$

Thus,

$$\mathbb{P}\left(f_m(X) \geq \mathbb{E} f_m(X) + t \right) \leq e^{-t^2/2} \quad \forall\, m$$

The claim follows by taking limit as $m \to \infty$ and noting that

$$\mathbb{E}\left[f_m(X)\right] = \mathbb{E}\left[(f * g_m)(X) \right]$$

$$= \mathbb{E}\left[f(X - Y_m) \right]$$

where $Y_m \sim N\left(0, \frac{I}{m}\right)$ is indep. of X, whereby

$$\lim_{m \to \infty} \mathbb{E}\left[f_m(X)\right] = \mathbb{E}\left[f(X)\right]. \qquad \blacksquare$$

$\boxed{D}$ Application: Concentration of supremum of Gaussian Processes

Lemma. (the finite index case)

Let $X_1, \ldots, X_n$ be jointly Gaussian with zero mean, i.e., $X_1, \ldots, X_n \sim N(0, \Sigma)$. Let $Z = \max\limits_{1 \leq i \leq n} X_i$.

Denote $\sigma^2 = \max\limits_{1 \leq i \leq n} \mathrm{Var}(X_i)$. Then,

$$\mathrm{Var}(Z) \le \sigma^2$$

and
$$\mathbb{P}(Z - \mathbb{E}Z > t) \le e^{-t^2/2\sigma^2}$$
$$\mathbb{P}(Z - \mathbb{E}Z < -t) \le e^{-t^2/2\sigma^2}.$$

Proof. Let A be a p.s.d. matrix satisfying $AA^T = \Sigma$. Then, $X = AY$, where $Y \sim N(0, I)$. Thus, $Z = f(Y)$ for $f(y) = \max_{1\le i\le n} (Ay)_i$. Then,

$$\begin{aligned}
|f(x) - f(y)| &\le \max_{1\le i\le n} |(Ax)_i - (Ay)_i| \\
&= \max_{1\le i\le n} \left| \sum_{j=1}^n A_{ij}(x_j - y_j) \right| \\
&\le \max_{1\le i\le n} \sqrt{\sum_{j=1}^n A_{ij}^2}\, \|x-y\| \\
&= \max_{1\le i\le n} \Sigma_{ii} \|x-y\| \le \sigma \|x-y\|.
\end{aligned}$$

The claim follows by the Gaussian Poincaré ineq. and the Gaussian conc. bound. ∎

We now consider a general Gaussian process $\{X_t\}_{t\in\mathcal{T}}$ where the index set $\mathcal{T}$ is a bounded, separable metric space. Further, the sample path X_t is assumed to be a.s. continuous. Therefore, $Z = \sup_{t\in\mathcal{T}} X_t = \sup_{t\in D} X_t$, where D is a countable dense subset of $\mathcal{T}$.

Therefore, $Z = \lim_{m \to \infty} Z_m$, a.s.

where $Z_m = \max_{1 \le i \le m} X_{t_i}$.

Let $\sigma^2 = \sup_{t \in \mathcal{T}} \mathrm{Var}(X_t) < \infty$. Then, using the Lemma above:

$$\mathbb{P}(Z_m \geq \mathbb{E}[Z_m] + t) \leq e^{-t^2/2\sigma^2},$$

$$\mathbb{P}(Z_m \leq \mathbb{E}[Z_m] - t) \leq e^{-t^2/2\sigma^2}$$

Since $Z_m \uparrow Z$ a.s., the proof will follow from monotone convergence theorem.

Applications of BLS and GLS to be seen later:

Threshold phenomena, analysis of LASSO