Concentration inequalities: Beyond the sub-Gaussian case [& Johnson-Lindenstrauss for finite sets]

So far we saw concentration results about sub-Gaussian tails.

Sub-Gaussian variables

For r.v. Z, define the log moment generating fn. (or cumulant gen. fn.)

$$\psi_Z(\lambda) := \log E[\exp(\lambda Z)]$$

Assume $EZ = 0$.

Then, Z is said to be sub-Gaussian with parameter ν, denoted $Z \in \mathcal{G}(\nu)$ if

$$\psi_Z(\lambda) \leq \frac{\nu\lambda^2}{2} \quad \forall \lambda \in \mathbb{R}.$$

Examples:-

(1) Bdd. variables: if $EZ=0$ & $Z \in [a,b]$ w.p. 1, then

$$Z \in \mathcal{G}\left(\frac{(b-a)^2}{4}\right) \quad \text{[Hoeffding's Lemma]}$$

(2) Gaussian:- if $Z \sim N(0,\sigma^2)$, then

$$Z \in \mathcal{G}(\sigma^2).$$

Pf:- $$E[\exp(\lambda Z)] = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} \exp\left(\lambda x - \frac{x^2}{2\sigma^2}\right)dx = \frac{\exp\left(\frac{\lambda^2\sigma^2}{2}\right)}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp\left(-\frac{(x-\lambda\sigma^2)^2}{2\sigma^2}\right)dx$$
$$= \exp\left(\frac{\lambda^2\sigma^2}{2}\right).$$

Main property: (1) If $X_1, X_2, \dots X_n$ are independent, $X_i \in \mathcal{G}(\nu_i)$, $E[X_i]=0$

then $$\sum_{i=1}^n X_i \in \mathcal{G}\left(\sum_{i=1}^n \nu_i^2\right).$$

(2) If $X \in \mathcal{G}(\nu)$, then $\forall\ t > 0$,

$$\max\left(P_r(X > t),\ P_r(X < -t)\right) \leq \exp\left(-\frac{t^2}{2\nu}\right).$$

This led, for example, to the following

Thm :- If $(X_i)_{i=1}^n$ are independent random variables taking values in the intervals $([a_i, b_i])_{i=1}^n$, then

$$\max\left(\Pr[X - E[X] > t],\ \Pr[X - E[X] < -t] \right) \leq \exp\left(-\frac{2t^2}{\sum (b_i - a_i)^2}\right). \quad \left[\begin{array}{c}\text{Hoeffding} \\ \text{inequality}\end{array}\right]$$

The dimension reduction problem

Input :- n points in a metric space (w.log, in $\mathbb{R}^n$ (why?))

Goal :- Project the points to $O(\log n)$ dimensions so that all pairwise distances are approximately preserved.

The Johnson-Lindenstrauss construction

$A = \frac{1}{\sqrt{d}}(A_{ij})_{d \times n}$, each A_{ij} is an independent random variable with $E[A_{ij}^2] = 1$, $E[A_{ij}] = 0$
$A_{ij} \in \mathcal{G}(1)$.

Thm : If $S \subseteq \mathbb{R}^n$ is of size n and $d \geq \frac{100v}{\varepsilon^2} \log\left(\frac{n}{\sqrt{\delta}}\right)$, then

$$1 - \varepsilon \leq \frac{\|Ax - Ay\|_2}{\|x - y\|_2} \leq 1 + \varepsilon \quad \forall\, x \neq y \in S.$$

Proof :- Let u be any unit vector in $\mathbb{R}^n$.

Then $\forall i \in [d]$, $(Au)_i = \frac{1}{\sqrt{d}} \sum A_{ij} u_j$ satisfies.

$$E[(Au)_i] = 0 \qquad E[(Au)_i^2] = \frac{1}{d} \sum_j u_j^2 E[A_{ij}^2] = \frac{1}{d}.$$

$$\Rightarrow E[\|Au\|_2^2] = \sum_{i=1}^{d} \frac{1}{d} = 1.$$

Further,

$$(Au)_i \in \mathcal{G}\left(\frac{1}{d} \sum_{j=1}^{n} u_j^2\right) = \mathcal{G}\left(\frac{1}{d}\right).$$

We are interested in the random variable

$$Z = \|Au\|_2^2 = \sum_{i=1}^{d} (Au)_i^2$$

But we only know that Au_i is $\mathcal{G}\left(\frac{1}{d}\right)$; not $(Au)_i^2$.

Thm:- (Special case of Bernstein's inequality) If $Z \sim \mathcal{G}(\nu)$ then

$$\psi_{Z^2 - E[Z^2]}(\lambda) \leq \frac{\alpha\lambda^2}{2(1-c\lambda)}, \quad \forall \lambda : |\lambda| < \frac{1}{c}$$

where $\alpha := 16\nu^2$, $c = 2\nu$.

Pf:- We first note

$$\psi_{Z^2 - E(Z^2)}(\lambda) = \log E\left[\exp\left(\lambda\left(Z^2 - E[Z^2]\right)\right)\right]$$

$$= \log E\left[e^{\lambda Z^2}\right] - E\left[\lambda Z^2\right]$$

$$\leq E\left[e^{\lambda Z^2}\right] - 1 - E\left(\lambda Z^2\right) \quad \left(\because \log x \leq x - 1 \;\; \forall x > 0\right)$$

$$= E\left[\sum_{q=2}^{\infty} \frac{\lambda^q Z^{2q}}{q!}\right]$$

$$= \sum_{q=2}^{\infty} \frac{\lambda^q E[Z^{2q}]}{q!}.$$

[This steps require a limit interchange argument but can be justified via dominated convergence.]

Now, $$E[Z^{2q}] = \int_0^{\infty} \Pr[Z^{2q} > x]\, dx = 2q \int_0^{\infty} y^{2q-1} \Pr[|Z| > x]\, dy$$

$$\leq 4q \int_0^{\infty} y^{2q-1} \exp\left(-\frac{y^2}{2\nu}\right) dy \quad \left(\because Z \in \mathcal{G}(\nu)\right)$$

$\left(t = \frac{y^2}{2\nu}\right)$

$$= 2q \cdot (2\nu)^q \int_0^{\infty} t^{q-1} \exp(-t)\, dt = 2(q!)(2\nu)^q$$

$$\Rightarrow \quad \psi_{Z^2 - E[Z^2]}(\lambda) \leq 2\sum_{q=2}^{\infty} |2\lambda\nu|^q$$

$$= \frac{8\lambda^2\nu^2}{1-2|\lambda|\nu} = \frac{\alpha\lambda^2}{1-c|\lambda|}, \text{ where } \begin{array}{l}\alpha := 16\nu^2 \\ c := 2\nu, \text{ when} \\ |\lambda| < 1/c .\end{array}$$

Note: If r.v. Z satisfies $\psi_Z(\lambda) \leq \frac{\alpha\lambda^2}{1-c|\lambda|}$ $\forall |\lambda| < 1/c$.
it is known as a sub-Gamma r.v. with params. (α, c).
(denoted $Z \in \Gamma(\alpha, c)$)

Further, If $Z_1, Z_2, \ldots Z_n$ are independent and $Z_i \in \Gamma(\alpha_i, c)$, then,

$$\sum Z_i \in \Gamma\left(\sum \alpha_i, c\right).$$

Concentration for sub-Gamma variables

If $Z \in \Gamma(\alpha, c)$,

$$\Pr(Z > t) \leq \inf_{\lambda \in (0, 1/c)} \exp\left(-\lambda t + \frac{\alpha\lambda^2}{1-c\lambda}\right).$$

After some calculations, we find,

$$\Pr\left(Z > \sqrt{2\alpha t} + ct\right) \leq \exp(-t)$$

[Compare subGaussian case: If $Z \in \mathcal{G}(\nu)$
$\Pr\left[Z > \sqrt{2\nu t}\right] \leq \exp(-t)$.]

Similarly, because $\psi_{-Z}(\lambda) = \psi_Z(-\lambda) \leq \frac{\alpha\lambda^2}{1-c\lambda}$ for every $\lambda \in (0, 1/c)$,

we get $\Pr\left(Z < -\sqrt{2\alpha t} - ct\right) \leq \exp(-t)$

Together, $\Pr\left(|Z| > \sqrt{2\alpha t} + ct\right) \leq 2\exp(-t)$

Thus, in our setting, we have

$$((Au)_i)^2 \in \Gamma\left(\frac{16}{d^2}, \frac{2}{d}\right)$$

$$\|Au\|^2 \in \Gamma\left(\frac{16}{d}, \frac{2}{d}\right), \quad \text{and hence}$$

$$\Pr\left[\left|\|Au\|^2 - 1\right| > \sqrt{\frac{32t}{d}} + \frac{2t}{d}\right] \leq 2\exp(-t)$$

Put $t = \log\left(\frac{n^2}{\delta}\right)$. Then,

$$\Pr\left[\left|\|Au\|^2 - 1\right| > 4\sqrt{\frac{2t}{d}} + \frac{2t}{d}\right] \leq \frac{2\delta}{n^2}.$$

Let $d = \lceil 2\gamma^2 \varepsilon^{-2} t \rceil$ for some γ to be determined later. Then,

$$\Pr\left[\left|\|Au\|_2^2 - 1\right| > \frac{4\varepsilon}{\gamma} + \frac{\varepsilon^2}{\gamma^2}\right] \leq \frac{2\delta}{n^2}.$$

$\left[\frac{4}{\gamma} + \frac{\varepsilon}{\gamma^2} < 1 \right.$

or, $\frac{4\gamma+\varepsilon}{\gamma^2} \leq 1$

$\gamma^2 \geq 4\gamma + 1$

If $\gamma = 5$, we have

$$\Pr\left[\left|\|Au\|_2^2 - 1\right| > \varepsilon\right] \leq \frac{2\delta}{n^2} \quad \forall \text{ unit vector } v.$$

Applying this to the $\binom{n}{2}$ unit vectors $\left\{\frac{x-y}{|x-y|} \,\middle|\, x \neq y \in S\right\}$ and using a union bound gives the result.

Isoperimetry and concentration

Let $X = (X_1, X_2, \dots, X_n) \sim N(0, I_n)$.

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt \quad ; \quad \phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2).$$

$$\Pr[X \leq x] = \Phi(x) \quad \forall x.$$

$$\gamma(x) = 1/\left(\Phi^{-1}(x)\right)' = \phi\left(\Phi^{-1}(x)\right).$$

Isoperimetry: Consider a A and define

$$A_t = \bigcup_{x \in A} B_t(x).$$

"Surface area" : $\limsup_{t \downarrow 0} \frac{\mathrm{Vol}(A_t \setminus A)}{t} = \lim_{t \downarrow 0} \frac{P(A_t \setminus A)}{t}$

Gaussian isoperimetry: Among all sets A of a given volume, this quantity is minimized by halfspaces.

More precisely,

$$\lim_{t \downarrow 0} \frac{P(A_t \setminus A)}{t} \geq \gamma(P(A))$$

with equality when A is a halfspace.

[We will not prove this.]

Corollary: "Gaussian concentration"

$$P(A_t) \geq \Phi\left(\Phi^{-1}(A) + t\right).$$

<u>Proof</u>:- Assume $f(t) = \Phi^{-1}(P(A_t))$ is differentiable.

Then

$$f'(t) = \frac{1}{\gamma(P(A_t))} \cdot \frac{dP(A_t)}{dt} \underbrace{\geq 1}_{\text{by isoperimetry, since}}.$$

$$\frac{dP(A_t)}{dt} = \lim_{s \downarrow 0} \frac{P(A_{t+s} \setminus A_t)}{s}.$$

Thus,

$$\Phi^{-1}(P(A_t)) \geq \Phi^{-1}(A) + t$$

$$\Rightarrow \quad P(A_t) \geq \Phi\left(\Phi^{-1}(A_t) + t\right).$$

This is in fact equivalent to isoperimetry, since

$$\lim_{t \downarrow 0} \frac{P(A_t) - P(A)}{t} \geq \lim_{t \downarrow 0} \frac{\Phi(\Phi^{-1}(P(A_t)) + t) - \Phi(\Phi^{-1}(P(A_t)))}{t}$$

$$= \phi\left(\Phi^{-1}(P(A_t))\right) = \gamma(P(A_t)).$$

<u>Corollary</u>:- (Gaussian concentration). If f is 1-Lipschitz on $\mathbb{R}^n$

$$|f(x) - f(y)| \leq \|x - y\|_2 \quad \text{then, if } X \sim N(0, I_n)$$

$$P[f(X) - \text{Median} f(X) > t] \leq 1 - \Phi(t), \forall t > 0.$$

<u>Note</u>:- When $t > 0$,

$$1 - \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty \exp\left(-\frac{x^2}{2}\right) dx$$

$$\leq \frac{1}{t\sqrt{2\pi}} \int_t^\infty x \exp\left(-\frac{x^2}{2}\right) dx$$

$$= \frac{\exp(-t^2/2)}{\sqrt{2\pi t^2}}.$$

In the next few lectures, we shall prove the slightly weaker Borell-Tsirelson-Ibragimov-Sudakov inequality:

If $X \sim N(0, I_n)$

$$\Pr\left[f(X) - E[f(X)] > t \right] \leq \exp\left(-\frac{x^2}{2}\right)$$

when f is 1-Lipschitz.

Entropy and the Herbst argument

We now look at some of the above proofs more closely.

Consider for example the Hoeffding lemma.

Assume $E[Z]=0$, and consider

$$\psi_Z(\lambda) = \log E[\exp(\lambda Z)]$$

Then
$$\psi_Z'(\lambda) = \frac{E[Z\exp(\lambda Z)]}{E[\exp(\lambda Z)]}$$

$$\psi_Z''(\lambda) = \frac{E[Z^2\exp(\lambda Z)]E[\exp(\lambda Z)] - E[Z\exp(\lambda Z)]^2}{E[\exp(\lambda Z)]^2}$$

$= \mathrm{Var}_{P'}[Z]$, where P' is the probability distribution given by

$$E_{P'}[Y] = \frac{E[Y\exp(\lambda Z)]}{E[\exp(\lambda Z)]} \quad \forall Y.$$