Día 3: Desigualdades sin independencia

Hasta ahora hemos visto concentración dada por la suma de variables independientes, en esta clase veremos que la concentración puede surgir en más contextos.

Vamos a estudiar funciones $f : \mathbb{S}^{n-1} \to \mathbb{R}$ y nos interesa entender cuando

$$f(X) \approx \mathbb{E} f(X)$$

$\uparrow$ $\mathrm{Unif}(\mathbb{S}^{d-1})$

Este tipo de concentración se da cuando la función f es relativamente estable.

Teorema: Considere un vector aleatorio $X \sim \mathrm{Unif}(\sqrt{n}\, \mathbb{S}^{n-1})$. Consideramos una función L-Lipschitz $f : \sqrt{n}\, \mathbb{S}^{n-1} \to \mathbb{R}$. Entonces,

$$\| f(X) - \mathbb{E} f(X) \|_{\psi_2} \leq C L.$$

$\dashv$

Corolario: Asuma $X \sim \mathrm{Unif}(\mathbb{S}^{n-1})$ y $f : \mathbb{S}^{n-1} \to \mathbb{R}$ es L-Lipschitz. Entonces

$$\| f(X) - \mathbb{E} f(X) \|_{\psi_2} \leq \frac{cL}{\sqrt{n}}.$$

Equivalentemente,

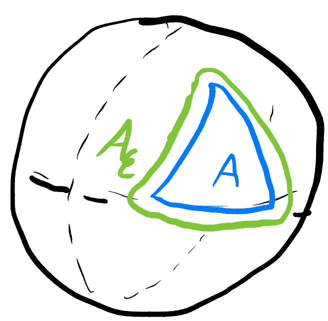
$$\mathbb{P}(|f(X) - \mathbb{E} f(X)| > t) \leq 2 \exp\left(- \frac{cnt^2}{L} \right).$$

Ejercicio: Probar el corolario.

Para probar este teorema usaremos un resultado profundo conocido como desigualdad isoperimétrica.

Dado un conjunto $A \subseteq \mathbb{S}^{n-1}$, define

$$A_\varepsilon = \{ x \in \mathbb{S}^{n-1} \mid \exists y \in A \text{ t.q. } \|x - y\| \leq \varepsilon \}$$



Teorema (*) Entre todos los conjuntos $A \subseteq \mathbb{S}^{n-1}$ con un area dada $\text{Area}(A)$ los casquetes esfericos (vecindarios de un solo punto) minimizan $\text{Area}(A_\varepsilon)$ para cualquier $\varepsilon > 0$ fijo.

Este resultado no lo probaremos.

Pero lo usaremos para probar el siguiente resultado.

Lemma: Sea $A \subseteq \sqrt{n}\,\mathbb{S}^{n-1}$ un conjunto. Defina $\sigma(A) = \text{Area}(A)/\text{Area}(\sqrt{n}\,\mathbb{S}^{n-1})$. Entonces si $\sigma(A) \geq 1/2$, para todo $t \geq 0$

$$\sigma(A_t) \geq 1 - 2\exp(-ct^2).$$

Prueba: Consideremos

$$H := \{ x \in \sqrt{n}\,\mathbb{S}^{n-1} : x_1 \leq 0 \}.$$

Por nuestras hipotesis, $\sigma(A) \geq \sigma(H)$, luego por la desigualdad isoperimétrica

Ejercicio $\sigma(A_t) \geq \sigma(H_t)$

Hecho: Las componentes de $X \sim \text{Unif}(\sqrt{n}\,\mathbb{S}^{n-1})$ satisfacen $\|X_i\|_{\psi} \leq C$. $\dashv$

Supongamos que $x \in \sqrt{n}\,\mathbb{S}^{n-1}$ satisface $x_1 \leq t/2$

$$\Rightarrow \quad y_j = \begin{cases} x_j & j \neq 1 \\ -x_1 & j = 1 \end{cases} \quad \Rightarrow \quad y \in H$$

$$\|x - y\| = 2x_1 \leq t$$

$$\Rightarrow \{x \in \mathbb{S}^{d-1} : x_1 \leq t/2\} \subseteq H_t.$$

Entonces

$$\sigma(H_t) \geq \mathbb{P}(X_1 \leq t/2) \geq 1 - 2\exp(ct^2).$$

$\square$

Consecuencia

En altas dimensiones, un slice en el ecuador es casi igual de pesada que toda la bola.

Prueba del Teorema (✱)

SPDG asumimos $L = 1$. Sea M la media de $f(X)$, es decir

$$\mathbb{P}(f(X) \leq M) \geq 1/2 \quad y \quad \mathbb{P}(f(X) \geq M) \geq 1/2.$$

Defina el conjunto de sub nivel

$$A = \{X \in \sqrt{n}\,\mathbb{S}^{n-1} : f(X) \leq M\}.$$

Entonces $\mathbb{P}(A) \geq 1/2$ y por el lema anterior

$$\mathbb{P}(A_t) \geq 1 - 2\exp(-ct^2).$$

Por el otro lado si $x \in A_t \Rightarrow \exists y \in A$ tq $\|x-y\| \leq t$. Luego

$$f(x) \leq f(y) + \|x-y\| \leq M+t$$

1-Lipschitz

Entonces

$$1-2\exp(-ct^2) \leq P(A_t) \leq P(f(x) \leq M+t)$$

Repitiendo el argumento con $-f$ obtenemos

$$\mathbb{P}(f(x)-M \geq -t) \geq 1-2\exp(-ct^2).$$

Combinando estos dos resultados obtenemos que

$$\|f(x)-M\|_{\psi_2} \leq C.$$

Casi tenemos el resultado que queremos, pero tenemos M en vez de $\mathbb{E}f(x)$.

Lema (*): Si W es una v.a. con $\|W\|_{\psi_2} \leq 1$. Entonces

$$\|W-\mathbb{E}W\|_{\psi_2} \leq C.$$

$\dashv$

Con este Lema si $f(X) - M = W$

$$\Rightarrow \| f(X) - M - \mathbb{E}f(X) + M \|_{\psi_2} \leq C.$$

Lo que prueba el teorema.

Finalmente tenemos que probar Lema ($\Diamond$).

Prueba: Por desigualdad triangular

$$\|X - \mathbb{E}X\|_{\psi_2} \leq \|X\|_{\psi_2} + \|\mathbb{E}X\|_{\psi_2}.$$

Recordemos que $\|1\|_{\psi_2} \leq c$. Luego

$$\begin{aligned}\|\mathbb{E}X\|_{\psi_2} &= |\mathbb{E}X| \, \|1\|_{\psi_2} \\ &\leq c \, \mathbb{E}|X| \\ &= c \, \|X\|_1 \\ &\leq c \, \|X\|_{\psi_2}.\end{aligned}$$

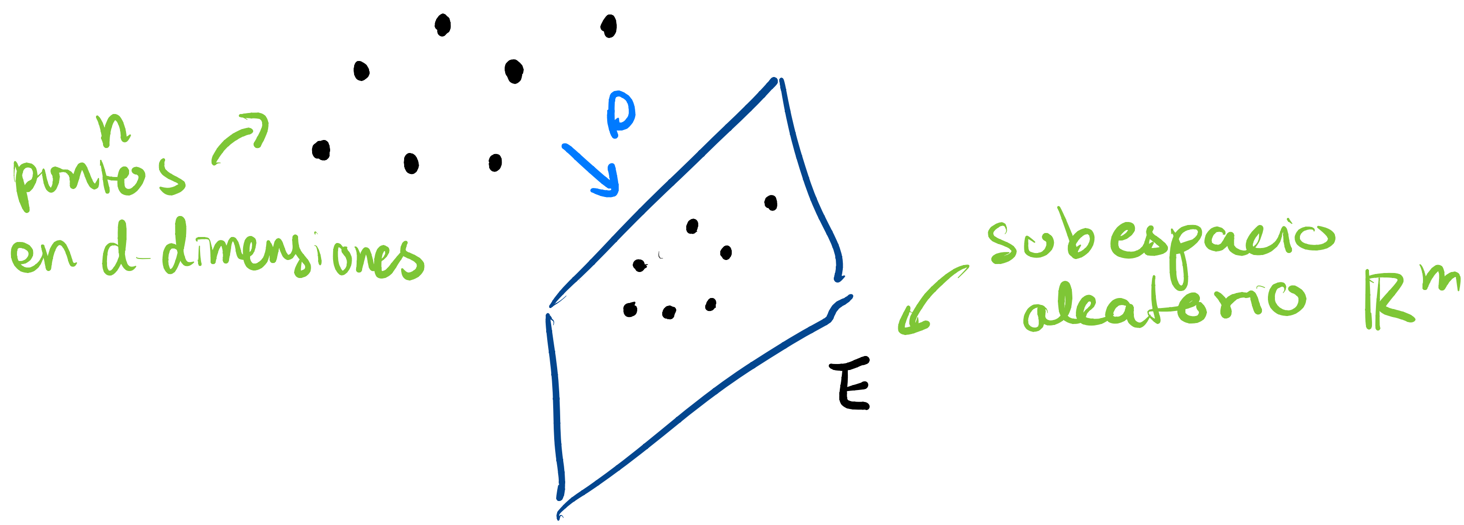
Caracterización de $\|.\|_{\psi_2}$ con momentos.

$\square$

$\square$

Aplicación reducción de dimensión

Recordemos la aplicación del primer día



El espacio de subespacios de dim m en $\mathbb{R}^d$, $G_{d,m}$, forma un grupo compacto y por consiguiente tiene una medida de Haar.

Johnson – Lindenstrauss Lema.

Teorema Sea $S \subseteq \mathbb{R}^d$ con $|S| = n$. Fije $\varepsilon > 0$ y asuma

$$m \geq \frac{C}{\varepsilon^2} \log n.$$

Considere $E \sim \mathrm{Unif}(G_{d,m})$. Sea P la proyección ortogonal a E. Entonces con probabilidad por lo menos $1 - 2\exp(-c\varepsilon^2 m)$, La proyección escalada

$$Q = \sqrt{\frac{d}{m}}\, P$$

satisface qué

$$(1-\varepsilon)\|x-y\|_2 \le \|Qx - Qy\|_2 \le (1+\varepsilon)\|x-y\|_2.$$

$\dashv$

Para probar este resultado usaremos el siguiente lema.

Lema: Sea $z \in \mathbb{R}^d$ un punto fijo. Entonces

(i) $$\left(\mathbb{E}\|Pz\|_2^2\right)^{1/2} = \sqrt{\frac{m}{d}}\,\|z\|_2.$$

(ii) Con probabilidad por lo menos $1 - 2\exp(-c\varepsilon^2 m)$ tenemos

$$(1-\varepsilon)\sqrt{\frac{m}{d}}\,\|z\|_2 \le \|Pz_2\|_2 \le (1+\varepsilon)\sqrt{\frac{m}{d}}\,\|z\|_2.$$

$\dashv$

Prueba del Lema: SPDG $\|z\|_2 = 1$.

Suponga que $E_0 = \{x : x_i = 0$ últimas $d-m$ comp.$\}$

Note que si R es una rotación aleatoria uniforme

Basis for E_0

$$E \overset{(d)}{=} RE_0 \qquad P \overset{(d)}{=} RU_0U_0^TR^T$$

$\sim \mathrm{Unif}(\mathbb{S}^{d-1})$

Entonces

$$\|Pz\| = \|RU_0U_0^TR^Tz\| = \|U_0U_0^TRz\|_2$$

Sin perdida de generalidad $E = E_0$ y $z \sim \mathrm{Unif}(\mathbb{S}^{d-1})$.

Note que

$$Pz = \begin{pmatrix} z_1 \\ \vdots \\ z_m \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad y \quad \mathbb{E}\|z\|^2 = 1$$

Como los z_i's están identicamente distribuidos

$$\mathbb{E}\, z_i^2 = \frac{1}{d} \qquad \forall i$$

Luego

$$\mathbb{E}\|Pz\|^2 = \frac{m}{d}.$$

Esto prueba (i). Para probar (ii) note que la función $f : \mathbb{S}^{d-1} \to \mathbb{R}$

$$z \longmapsto \|Pz\|$$

Chequear

es 1-Lipschitz. Luego tenemos que

$$\big\| \|Pz\|_2 - \mathbb{E}\|Pz\|_2 \big\|_{\psi_2} \leq C/\sqrt{d}.$$

Entonces

Lemita: Suponga que W es una v.a t.q. $\|W\|_{\psi_2} < \infty$. Entonces

$$\|W - \|W\|_2\|_{\psi_2} \leq C \|W\|_{\psi_2}$$

Prueba:

$$\|W - \|W\|_2\|_{\psi_2} \leq \|W\|_{\psi_2} + \|\|W\|_2\|_{\psi_2}$$

$$\leq \|W\|_{\psi_2} + \|W\|_2 \|1\|_{\psi_2}$$

Caracterización de momentos.

$$\leq \|W\|_{\psi_2} + \sqrt{2}C\|W\|_{\psi_2}.$$

Con el lemita concluimos que

$$\mathbb{P}\left(\left|\|Pz\|_2 - \sqrt{\tfrac{m}{d}}\right| \geq t\right) \leq 2\exp(-ct^2)$$

Lo cual es equivalente al enunciado del lema. $\square$

Con este lema estamos listos para probar Johnson-Lindenstrauss.

Consideremos el conjunto $T = \{x-y : x,y \in S,\ x \neq y\}$. Usando el lema que acabamos de probar

$$\mathbb{P}(\exists z \in \mathcal{T} \text{ tq } z \text{ no satisface } (\;))$$

$$\leq \sum_{z \in \mathcal{T}} \mathbb{P}(z \text{ no satisface } (\;))$$

$$\leq 2|\mathcal{T}| \exp(-c_1 \varepsilon^2 m)$$

$$= 2 \binom{n}{2} \exp(-c_1 \varepsilon^2 m)$$

$$\leq 2 n^2 \exp(-c_1 \varepsilon^2 m)$$

$$= 2 \exp(2 \log(n) - c_1 \varepsilon^2 m)$$

Como $m \geq \frac{C}{\varepsilon^2} \log(n)$, podemos asignar $C = \frac{4}{c_1}$ para obtener

$$\mathbb{P}(\exists z \in \mathcal{T} \text{ tq } z \text{ no satisface } (\;))$$

$$\leq 2 \exp\left(-\frac{c_1}{2} \varepsilon^2 m\right).$$

Lo que establece el resultado.

$\square$

Recapitulando

En estas clases cubrimos:

- Dos ejemplos
- Desigualdades clásicas
- Matrices aleatorias
- ε- Mallas
- Norma de matrices sub-gaussianas
- Desigualdades isoperimétricas
- Concentración en la esfera
- Johnson-Linderstrauss

Día 1

Día 2

Día 3

Hay muchísimo tema que no cubrimos, pero espero que la clase haya servido como un abre bocas.

Un par de referencias buenas:

- Vershynin.
- Tropp.