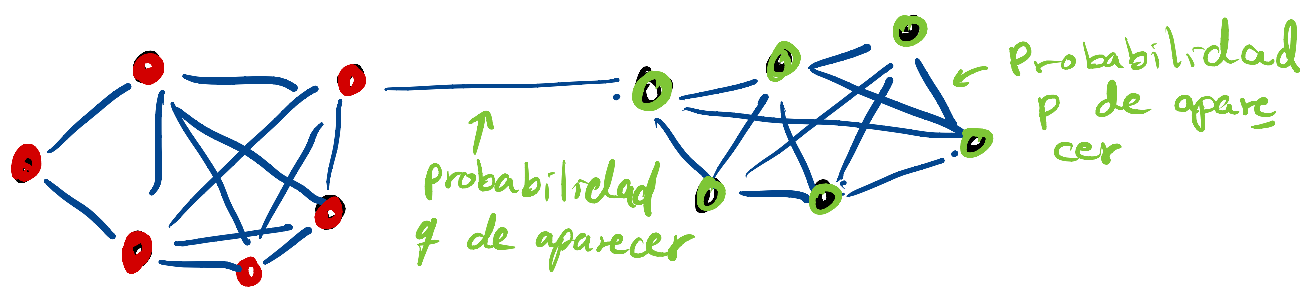
Día 1 (7/Nov)

El mundo es aleatorio, lo que puede ser difícil de navegar. Pero muchas veces las cantidades aleatorias parecen más bien deterministas y se pueden manejar.

Ejemplo 1: Detección de comunidades

Asumamos que tenemos un grafo aleatorio G con dos comunidades de nodos



La mitad de los nodos son rojos y la otra mitad verdes (nosotros no vemos los colores), y el grafo se genera

$$\forall\, i,j \in V \quad \mathbb{P}((i,v) \in V) = \begin{cases} p & \text{si color}(i) = \text{color}(j) \\ q & \text{De lo contrario} \end{cases}$$

con $p > q$.

Nuestro objetivo es estimar los colores.
Sea A_G la matriz de adyacencia. Algo razonable es ver que pasa en promedio:

$$\mathbb{E} A_G = \left[\begin{array}{ccc|ccc} q & \cdots & q & p & \cdots & p \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ q & \cdots & q & p & \cdots & p \\ \hline p & \cdots & p & q & \cdots & q \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p & \cdots & p & q & \cdots & q \end{array}\right]$$

Hecho: Los valores/vectores propios de $\mathbb{E} A_G$ son

$$\lambda_1 = \left(\frac{p+q}{2}\right)n, \quad u_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \lambda_2 = \left(\frac{p-q}{2}\right)n, \quad u_2 = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \hline -1 \\ \vdots \\ -1 \end{bmatrix}$$

Este tipo nos da toda la info.

Idea: Calular u_2 de A_G y usarlo como una aproximación de u_2 de $\mathbb{E}A_G$.

¿Qué tan bien funciona esto?
Muy bien cuando $p-q$ es grande. PLOT

¿Por qué?
La matriz se concentra $\frac{\|A_G - \mathbb{E}A_G\|}{\|\mathbb{E}A_G\|} \to 0$. PLOT

Ejemplo 2: Reduciendo la dimensión

Supongamos que tenemos n puntos $v_1, \dots, v_n \in \mathbb{R}^d$ y que nos interesa comprimirlos con un mapa

$$F : \mathbb{R}^d \to \mathbb{R}^m$$

(♡) con $m << d$ tal que $\forall\, i, j$

$$(1-\varepsilon)\|v_i - v_j\| \leq \|F(v_i) - F(v_j)\| \leq (1+\varepsilon)\|v_i - v_j\|$$

casi una isometría para estos puntos.

Idea: Tomemos P una matriz de proyección a un subespacio aleatorio dimensión m, definimos $F(x) = \sqrt{\frac{m}{d}}\, Px$.

¿Que pasa en promedio?

$$E\|F(u) - F(v)\|^2 = \|u - v\| \quad \forall\, u, v \in \mathbb{R}^d$$

El lema de Johnson-Lindenstrauss nos dice que si $m \geq \frac{c}{\varepsilon^2} \log n$, entonces (♡) se tiene con alta probabilidad.

PLOT.

Desigualdades de concentración clásicas.

Proposición (Desigualdad de Markov)

Suponga que X es una v.a. positiva. Para cualquier $t > 0$, tenemos

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}.$$

Prueba: Represente

$$X = X \mathbb{1}\{X \geq t\} + X \mathbb{1}\{X < t\}$$

Tomando promedios

$$\begin{aligned}
\mathbb{E}X &= \mathbb{E}[X \mathbb{1}\{X \geq t\}] + \mathbb{E}[X \mathbb{1}\{X < t\}] \\
&\geq t \, \mathbb{E} \mathbb{1}\{X \geq t\} \\
&= t \, \mathbb{P}(X \geq t).
\end{aligned}$$

$\square$

Lo que haremos es mejorar este resultado aplicando transformaciones a X.

Corolario (Chebychev) Sea X una v.a. con esperanza μ y varianza σ^2. Entonces, para todo $t > 0$, tenemos

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Ejercicio.

Def: Una v.a. X tiene distribución Bernoulli simétrica si

$$P(X=-1)=P(X=1)=1/2.$$

Teorema (Hoeffding) Sean $X_1,\dots,X_n$ v.a. ind. Bernoulli simétricas y sea $a=(a_1,\dots,a_n)\in\mathbb{R}^n$. Entonces, para cualquier $t\geq 0$, tenemos

$$\mathbb{P}\left(\sum a_i X_i \geq t\right) \leq \exp\left(-\frac{t^2}{2\|a\|_2^2}\right)$$

Si $a_i=\frac{1}{n}$
$\|a\|^2=\frac{1}{n}$

Prueba: La funcion $x\mapsto \exp(\lambda x)$ es creciente para todo λ, luego

$$(\star)\quad P\left(\sum a_i X_i\geq t\right)=\mathbb{P}\left(\exp\left(\lambda\sum a_i X_i\right)\geq \exp(\lambda t)\right)$$
$$\overset{\text{Markov}}{\leq} e^{-\lambda t}\underbrace{\mathbb{E}\exp\left(\lambda\sum a_i X_i\right)}_{\text{Función generadora de momentos.}}$$

Como los X_i son iid, tenemos

$$\mathbb{E}\exp\left(\lambda\sum^n a_i X_i\right)=\prod^n \mathbb{E}\exp(\lambda a_i X)$$

Hay dos posibles valores ± 1, entonces

$$\mathbb{E}\exp(\lambda a_i X) = \frac{1}{2}(\exp(\lambda a_i) + \exp(-\lambda a_i))$$
$$= \cosh(\lambda a_i)$$

Ejercicio (Taylor)

$$\leq \exp((\lambda a_i)^2/2)$$

Substituyendo en $(\star)$,

$$\mathbb{P}(\sum a_i X_i \geq t) \leq e^{-\lambda t} \prod \exp((\lambda a_i)^2/2)$$
$$= \exp(-\lambda t + \frac{\lambda^2}{2}\sum a_i^2)$$
$$= \exp(\underbrace{-\lambda t + \frac{\lambda^2}{2}\|a\|^2}_{\text{Cuadratica } q(\lambda)})$$

Minimizando $q'(\lambda) = 0 \iff \lambda = \dfrac{t}{\|a\|^2}$.

Substituyendo

$$\mathbb{P}(\sum a_i X_i \geq t) \leq \exp(-t^2/2).$$

$\square$

Ejercicio 2.2.7.

Observación: Esta estrategia de prueba se puede usar para probar desigualdades más generales (Chernoff).

Distribuciones sub-Gausianas

Algo natural es preguntarse para qué otras distribuciones tenemos la concentración dada por Hoeffding?

$$\mathbb{P}\left\{ \left| \sum_i a_i X_i \right| \geq t \right\} \leq 2 \exp\left(- \frac{c t^2}{\|a\|^2}\right).$$

Esto automaticamente nos da una restriccion:

$$\mathbb{P}\{ |X_i| > t \} \leq 2 \exp(-ct^2).$$

Llamaremos v.a. que satisfagan esta desigualdad, su-Gaussianas.

Ejemplos

- Gaussianas $\quad P(|X| \geq t) \leq 2\exp(-t^2/2)$
- Bernoulli $\quad$ (Hoeffding)
- Bounded $\quad$ (Ejercicio).

Proposición (Caracterización) Sea X una variable aleatoria. Las siguientes propiedades son equivalentes (modulo factores constantes entre los parametros K_i):

(i) Las colas de X satisfacen

$$\mathbb{P}(|X| > t) \leq 2\exp(-t^2/K_1).$$

(ii) Los momentos de X satisfacen

$$\|X\|_{L^p} = (\mathbb{E}|X|^p)^{1/p} \leq K_2\sqrt{p}.$$

(iii) La fonción generadora de momentos (FGM) de X^2 satisface

$$\mathbb{E}\exp(\lambda^2 X^2) \leq \exp(K_3^2\lambda^2) \quad \forall |\lambda| \leq \frac{1}{K_3}.$$

(iv) La FGM de X^2 satisface que

$$\mathbb{E}\exp(X^2/K_4^2) \leq 2$$

Si $\mathbb{E}X = 0$ entonces la siguiente también es equivalente a (i) - (iv)

(v) La FGM de X satisface

$$\mathbb{E}\exp(\lambda X) \leq \exp(K_5\lambda^2) \quad \forall \lambda \in \mathbb{R}.$$

Prueba:

(i) $\Rightarrow$ (ii). $\overbrace{\text{Sin perdida de generalidad}}^{\text{SPDG}}$ $K_1 = 1$.

Usando la identidad $\left(\mathbb{E}X = \int_0^\infty \mathbb{P}(X \geq u)\,du\right)$

$$\mathbb{E}|X|^p = \int_0^\infty \mathbb{P}(|X|^p \geq u)\, du$$

Cambio de variable $t^p \leftarrow u$

$$= \int_0^\infty \mathbb{P}(|X| \geq t)\, p t^{p-1}\, dt$$

$$\leq 2p \int_0^\infty e^{-t^2} t^{p-1}\, dt$$

Cambio de variable $s \leftarrow t^2$

$$= p \underbrace{\int_0^\infty e^{-s} s^{\frac{p}{2}-1}\, ds}_{\text{Función Gamma } \Gamma(p/2)}$$

$$= p\, \Gamma(p/2)$$

$$\leq 3p\, (p/2)^{p/2} \quad \leftarrow \Gamma(x) \leq 3x^x \;\; \forall x \geq \tfrac{1}{2}$$

El resultado se sigue tomando raíces $(\cdot)^{1/p}$ y notando que $x^{1/x} \leq 2 \;\; \forall x > 0$.

(ii) $\Rightarrow$ (iii) SPDG $K_2 = 1$. Por Taylor,

$$\mathbb{E} \exp(\lambda^2 X^2) = \mathbb{E}\left[1 + \sum_{p=1}^{\infty} \frac{(\lambda^2 X^2)^p}{p!}\right] = 1 + \sum_{p=1}^{\infty} \frac{\lambda^{2p}\, \mathbb{E}\, X^{2p}}{p!}$$

Usando la cota $\mathbb{E}[X^{2p}] \leq (2p)^p$ y la aproximación de Stirling da $p! > (p/e)^p$.

Sustituyendo

$$\mathbb{E} \exp(\lambda^2 X^2) \leq 1 + \sum_{p=1}^{\infty} \frac{(2\lambda^2 p)^p}{(p/e)^p} = \sum_{p=0}^{\infty} (2e\lambda^2)^p$$

Si $2e\lambda^2 < 1$.

$$= \frac{1}{1 - 2e\lambda^2}$$

$\frac{1}{1-x} \leq e^{2x}$ si $x \in [0,1]$. $\leq \exp(4e\lambda^2)$

La cota se tiene cuando $|\lambda| \leq \frac{1}{4\sqrt{e}}$.

$(iii) \Rightarrow (iv)$ Trivial.

$(iv) \Rightarrow (i)$ SPDG $K_4 = 1$. Entonces

$$
\begin{aligned}
\mathbb{P}(|X| \geq t) &= \mathbb{P}(e^{X^2} \geq e^{t^2}) \\
&\leq e^{-t^2}\, \mathbb{E}\, e^{X^2} \\
&\leq 2e^{-t^2}.
\end{aligned}
$$

(i) se cumple con $K_1 = 1$.

La equivalencia con (v) se deja como ejercicio. $(iii) \Rightarrow (v)$, $(v) \Rightarrow (i)$.

$\square$

Definición Una v.a. X es sub-Gaussiana si satisface, cualquiera de las 5 propiedades. Su norma sub-Gaussiana, $\|X\|_{\psi_2}$ se define como

$$\|X\|_{\psi_2} := \inf\{K > 0 : \mathbb{E}\exp(X^2/K^2) \leq 2\}.$$

Pruebe que esto es una norma.

$\dashv$

Gracias a la prosición, si $\|X\|_{\psi_2} < \infty$

(ö) $\mathbb{P}(X \geq t) \leq 2\exp(-ct^2/\|X\|_{\psi_2}^2) \quad \forall t \geq 0$,

(ü) $\|X\|_{L^p} = C\|X\|_{\psi_2}\sqrt{p} \quad \forall p \geq 1$,

(ẍ) Si $\mathbb{E}X = 0 \Rightarrow \mathbb{E}\exp(\lambda X) \leq \exp(c\lambda^2\|X\|_{\psi_2}^2) \quad \forall \lambda \in \mathbb{R}$.

La desigualdad de Hoeffding generalizada

Proposición (++) Sean $X_1, \dots, X_n$ v.a. ind., sub-gaussianas con esperanza cero. Entonces

$$\left\| \sum_{i=1}^{n} X_i \right\|^2 \leq C \sum \|X_i\|_\psi^2$$

para una constante absoluta $C>0$.

Prueba: Sea $\lambda \in \mathbb{R}$,

$$\begin{aligned}\mathbb{E} \exp(\lambda \textstyle\sum X_i) &\leq \prod \mathbb{E} \exp(\lambda X_i) && \text{(Independencia)} \\ &= \prod \exp(c \lambda^2 \|X_i\|_{\psi_2}^2) && (\divideontimes) \\ &\leq \exp\left(\lambda^2 \left(c \textstyle\sum \|X_i\|_{\psi_2}^2\right)\right).\end{aligned}$$

$\square$

Un corolario directo de esta prosición:

Teorema Sean $X_1, \dots, X_n$ v.a. ind. sub-gaussianas con $\mathbb{E} X_i = 0$ y $a \in \mathbb{R}^n \setminus \{0\}$. Entonces, para todo $t \geq 0$, tenemos

$$\mathbb{P}\left\{ \left| \sum_{i=1}^{n} a_i X_i \right| \geq t \right\} \leq 2 \exp\left(- \frac{ct^2}{K^2 \|a\|_2^2}\right),$$

donde $K = \max_i \|X_i\|_{\psi_2}$ $\dashv$