PROBABILIDAD EN DIMENSIÓN ALTA

Curso de PEDECIBA Matemática .
(Primer semestre de 2019)

Comienzo de curso
Semana del 18 de marzo | CMAT-IMERL

Docente

Diego Armentano (CMAT)

Objetivo del curso

La ciencia de datos está en pleno auge y es necesario tener investigadores con conocimientos sólidos para poder abordar los problemas que surgen en el área. En este curso, "Probabilidad en dimensión alta", se verán técnicas que han sido fundamentales para el desarrollo de la ciencia de datos, y que aún hoy inspiran nuevas soluciones para problemas actuales. (Ver abajo *Descripción del tema del curso* para más detalles.)

Público objetivo

Estudiantes de maestría y doctorado en matemática, y maestría en ingeniería matemática, e investigadores que tengan curiosidad por esta nueva área.

Conocimientos previos recomendados

Además de los cursos básicos de cálculo y álgebra, se recomienda el curso de Probabilidad y Topología. Se sugiere un curso de teoría de la medida, pero no es obligatorio.

Para ciertos tópicos que veremos, conocimientos en geometría Riemanniana y análisis funcional serán bienvenidos, pero para nada excluyentes.

Programa del curso

Este curso seguirá de cerca las notas del curso de Roman Vershynin en la Universidad de California (ver Referencias). También utilizaremos las notas de Ramon van Handel en la Universidad de Princeton.

El programa del curso es el siguiente:

- Motivaciónes (cubrimiento de conjuntos, concentración de la medida, supremo de procesos, universalidad, fenómenos de transición).

- Preliminares sobre variables aleatorias (v.a.)
- Concentranción de sumas de v.a.
- Vectores aleatorios en grandes dimensiones
- Matrices aleatorias
- Concentración sin independencia
- Formas cuadráticas, simetrización y contracciones
- Procesos estocásticos
- Grandes desvíos en matrices aleatorias
- Esparsidad en ciencia de datos

Referencias

- *High-Dimensional Probability "An Introduction with Applications in Data Science"*, Roman Vershynin, University of California, Irvin
- *Probability in High Dimension*, Ramon van Handel, Universidad de Princeton.

Descripción del tema del curso

Definir qué entendemos por "probabilidad en dimension alta" no es sencillo dado que atraviesa transversalmente muchas áreas diferentes (matrices y grafos aleatorios, algoritmos randomizados, probabilidad en espacios de banach, geometría asintótica, mecánica estadística, machine learning, teoría de la información, ciencia de datos, etc.). Aún la naturaleza de la "dimensión alta" puede ser muy diferente. Por ejemplo podemos estudiar fenómenos donde participan un gran número de variables aleatorias, o por otro lado fenómenos determinísitcos que ocurren en espacios de dimensión muy alta donde métodos probabilisticos son utilizados como técnicas para estudiarlos.

Veamos algunos ejemplos de lo anterior. El primero referido al estudio de objetos en un espacio euclideano de dimensión muy alta. El resto es el caso de fenómenos que involucran grandes números de variables aleatorias.

- *Cubrimiento de conjuntos*

  Hay un teorema debido a Carathedory que dice que todo punto de la envolvente convexa $\operatorname{conv}(C)$ de un conjunto $C \subset \mathbb{R}^n$ puede ser expresado como combinación lineal de a lo sumo $n+1$ puntos de C. (Esta cota no puede ser mejorada como puede observarse tomando el ejemplo de un símplice.) Cuando la dimensión n es muy grande, uno puede interesarse por un resultado aproximado de lo anterior, y que involucre sólo algunos puntos en la combinación lineal. Como veremos en la motivación del curso, con técnicas probabilísticas veremos que se puede probar el siguiente resultado.

  > Teorema: *Sea $C \subset \mathbb{R}^n$ con diámetro menor o igual a 1. Entonces, para cualquier $x \in conv(C)$, y cualquier $k \in \mathbb{N}$, se tiene que existen $x_1, \dots, x_k$ en C tales que*
  >
  > $$\left\| x - \frac{1}{k}\sum_{i=1}^{k} x_i \right\|_2 \leqslant \frac{1}{\sqrt{k}}$$

  En otras palabras, los "centro de masa" de finitos puntos de C alcanzan para aproximar muy bien cualquier punto de su envolvente convexa. Lo más destacable de lo anterior es que el resultado anterior es independiente de la dimensión n (que puede ser astronómica).

  Como veremos, y el lector puede intuir, este resultado tiene implicancias importantes sobre el "packing problem" de conjuntos.

- *Concentración y universalidad*

  Por la ley de los grandes números sabemos que si $X_1, X_2, \dots$ son variables aleatorias i.i.d., (con esperanza finita) entonces

  $$\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \longrightarrow 0 \quad \text{cuando} \quad n \to \infty.$$

  Una forma de reescribir lo anterior es considerar la función lineal $f : \mathbb{R}^n \to \mathbb{R}$, $f(x_1, \dots, x_n) = \frac{1}{n}(x_1 + \cdots + x_n)$. Luego se tiene que la v.a. $f(X_1, \dots, X_n)$ se concentra en el valor esperado de la misma. Sin embargo, el resultado anterior puede extenderse de la siguiente manera.

  > *Si $X_1, X_2, \dots$ son v.a. independientes (o con dependencia débil), entonces la v.a. $f(X_1, \dots, X_n)$ se concentra en su valor medio, siempre que f no sea muy sensible a las coordenadas x_i.*

  Este tipo de fenómenos se denominan de *concentración*

  En la situación inicial de v.a. independientes, por el teorema central del límite uno puede conocer el comportamiento de las fluctuaciones alrededor de la media. De hecho, en el caso independiente resulta que

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \approx \text{Gaussiano} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(G_i - \mathbb{E}G_i)$$

donde $G_1, G_2, \ldots$ son gaussianas independientes con igual esperanza y varianza que X_i.

Una de las implicancias más importante de este tipo de resultados es la robustes del mismo; i.e., cuando n es grande la distribución de las v.a. X_i es irrelevante. Este tipo de resultados se lo denomina como *universalidad*. De hecho, algo más fuerte ocurre por ejemplo en el caso de la esperanza:

> *Si $X_1, X_2, \ldots$ son v.a. independientes (o con dependencia débil), entonces la $\mathbb{E}(f(X_1,\ldots,X_n))$ no depende de la distribución inicial de X_i, siempre que f sea suficientemente suave.*

La universalidad en este caso es interesante por la robustez del problema al hacer crecer n. Pero por otro lado es una herramienta fundamental para poder hacer cálculos de esperanzas en casos que casi sería imposible realizarlos con exactitud. A modo de ejemplo, ocurre muy seguido en el caso de matrices aleatorias cuando n es grande. Hacer cálculos exactos es casi imposible en general, pero sí es realizable en el caso de matrices con coeficientes gaussianos.

- *Transición de fase*

  Otro fenómeno que ocurre en dimensiones altas, un poco distinto a los anteriores, es el caso de las *transiciones de fase*, muy conocido en mecánica estadística.

  Grosso modo, este fenómeno se observa cuando hay un parámetro común que modela cierto fenómeno, que hace que los efectos visibles cambien suavemente a medida que el parámetro se mueve. Sin embargo cuando la dimensión es muy grande puede ocurrir un cambio drástico en cierto valor singular del parámetro, fenómeno conocido como *cambio de fase*. Veamos un ejemplo sencillo de esto.

  Supongamos que tenemos v.a. $X_1, X_2, \ldots$, Bernoulli independientes de parámetro p, i.e.,

$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_1 = 0) = 1 - p.$$

  Fijado n, sea Y_n la v.a. que toma el valor 1 si el número de "unos" es mayor o igual al número de "ceros". Es decir,

$$Y_n = \mathbb{1}_{\{\frac{1}{n}\sum_{i=1}^{n} X_i \geqslant \frac{1}{2}\}}$$

  Un cálculo sencillo demuestra que cuando $n \to \infty$ se tiene

$$\mathbb{E}(Y_n) \longrightarrow \begin{cases} 0 & p < 1/2 \\ 1/2 & p = 1/2 \\ 1 & p > 1/2 \end{cases} \qquad \text{cuando} \quad n \to \infty.$$

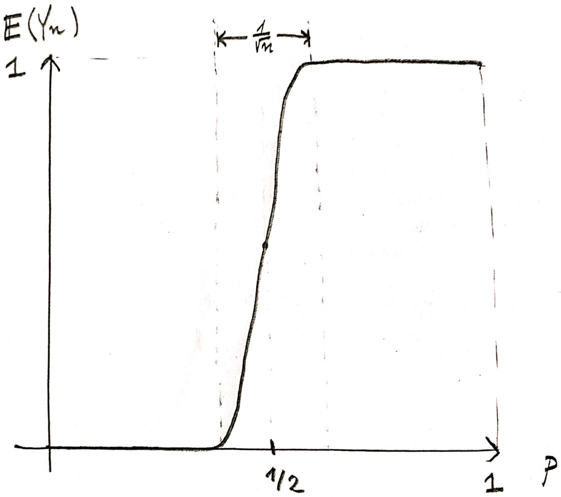

Figure 1: Esperanza de Y_n para n grande.

Analizando más en detalle, y utilizando el TCL para conocer que las fluctuaciones son del orden de $1/\sqrt{n}$, obtenemos que para n fijo la esperanza de Y_n tiene la siguiente representación gráfica (ver Figura).

Como se puede observar, $\mathbb{E}(Y_n)$ se mueve suvemente para valores de n chicos, pero en el límite se crea una singularidad en $p^* = 1/2$, lo que se denomina transición de fase. $\square$