# Curso de Maestría en Estadística

# Centro de Investigación en Matemáticas, A.C.

# Inferencia estadística para datos complejos

Agosto-diciembre 2016

**Profesor**: Miguel Nakamura (ext. 4539, oficina K–303, nakamura@cimat.mx). Para agendar citas, favor de ingresar a doodle.com/mnakamura.

**Horario de clase**: Por anunciarse.

**Sitio web para materiales diversos y avisos:**

https:\\xxx.xxx.xxx

# Objetivo y descripción

El objetivo general es incursionar en tendencias actuales para el análisis de datos, y ello incluye algunas de situaciones motivadas por la modernidad, en las cuales la estructura de datos dista mucho de ser la premisa canónica de observaciones finitas de vectores aleatorios de dimensión manejable—tanto conceptual como computacionalmente. Lo anterior incluye las situaciones en las que los datos son objetos complejos, o en las que su número es varios órdenes de magnitud mayor a lo que dictan los supuestos convencionales de estadística matemática.

Durante la celebración del Año Internacional de la Estadística 2013 en el CIMAT, se contó con el privilegio de recibir a distinguidos conferencistas invitados. En particular, E. Candes, J.S. Marron, y M. Jordan tuvieron en común presentar reseñas visionarias de retos y temas de frontera. Aprovechando este *leitmotiv*, así como la coyuntura mencionada, el eje conductor y la dinámica del curso serán no-convencionales: Consistirán de analizar algunos artículos insignia de estos autores, y que los detalles técnicos necesarios para entender esos artículos se conviertan de facto en el temario del curso. Como ejemplos, en virtud de que [4] tiene que ver con bootstrap para datos masivos, el tema de bootstrap introductorio y convencional será un candidato idóneo para abordarse; similarmente, entre otras cosas, [6] asume entre sus premisas algo de componentes principales y análisis de formas.

Por su concepción, el curso tendrá una dinámica muy adaptativa, inmersa en un ambiente intenso de aprendizaje colaborativo. La evaluación del curso consistirá de exposiciones en clase y tareas. Habrá mucha flexibilidad en los temas de dichas exposiciones, como para adaptarse a gustos e intereses de los participantes.

# Temario

1. Introducción a complejidad de datos y retos estadísticos.
2. Datos como curvas y análisis de datos funcionales.
3. Datos como formas.
4. Datos en el contexto de análisis de imágenes.
5. Datos en variedades.
6. Bootstrap.
7. Bootstrap para datos masivos.
8. Otras extensiones.

# Referencias Bibliográficas

1. Bühlmann, P. and van de Geer, S. (2011), *Statistics for High-Dimensional Data*, Springer-Verlag.
2. Candes, E., and Tao, T. (2007), “The Dantzig Selector: Statistical Estimation when *p* is much larger than *n*”, *The Annals of Statistics*, vol. 35, no. 6, 2313–2351.
3. Cui, X. and Wilson, J. (2008), “On the Probability of Correct Selection for Large *k* Populations, with Application to Microarray Data”, *Biometrical Journal*, vol. 50, no. 5, 870–883.
4. Dryden, I.L., and Mardia, K.V. (1998), *Statistical Shape Analysis*, Wiley.
5. Kleiner, A., Talwalkar, A., Sarkar, P. and Jordan, M.I. (2014), “A scalable bootstrap for massive data”, *Journal of the Royal Statistical Society, Series B*, doi:10.1111/rssb.12050
6. Hall, P., Marron, J.S., and Neeman, A. (2005), “Geometric representation of high dimension, low sample size data”, *Journal of the Royal Statistical Society, Series B*, vol. 67, Part 3, 427–444.
7. Marron, J.S. and Alonso, A.M. (2014), “An Overview of Object Oriented Data Analysis”, *Biometrical Journal*, vol. 56, no. 5, 732–753.
8. Ramsay, J.O. (2005), *Functional Data Analysis*, Wiley.