

Pablo Groisman

Universidad de Buenos Aires e IMAS-CONICET, Argentina

pgroisma@dm.uba.ar

Consideraremos la consistencia de las k -medias (de Fréchet) en espacios métricos cuando las distancias son desconocidas y deben estimarse a partir de los datos. Mostraremos que las k -medias empíricas son consistentes si los espacios métricos empíricos convergen al poblacional en el sentido medido de Gromov-Hausdorff (mGH). Como consecuencia, obtenemos (a) consistencia de las k -medias calculadas con distancias de Fermat, (b) consistencia de las k -medias calculadas con el estimador de distancia geodésica de Isomap y (c) consistencia del k -baricentro empírico de Wasserstein calculado con estimadores basados en muestras de las medidas involucradas. Para (a) y (b), estos resultados son nuevos incluso para el caso $k = 1$, hasta donde sabemos. Para (c), son bien conocidos para $k = 1$, pero no para $k > 1$, hasta donde sabemos. No asumimos la unicidad del baricentro, pero cuando éste es único, los resultados son más fuertes.

Trabajo en conjunto con Matthieu Jonckhere (LAAS, CNRS), Jordan Serres (Sorbonne) y Mariela Sued (Universidad de San Andrés).