

Séminaire de Probabilités et Statistique

Mardi 9 Avril à 14h00

Laboratoire Dieudonné
Salle de conférence - LJAD

Martin Royer

Université Paris-Sud

K-means corrigé, optimalité statistique et optimisation convexe .

On traite du problème du partitionnement ("clustering") d'un ensemble d'entités en K groupes à l'aide d'un modèle statistique simple et dans un cadre non-asymptotique : peut-on discerner des groupes dans ces entités de façon optimale ? En partant de l'étude de l'estimateur classique des K -moyennes ("K-means"), on donne de nombreux éléments de réponse pour les petites et grandes dimensions grâce à une reformulation puissante par l'optimisation convexe [1], qui éclaire par ailleurs notre compréhension d'autres estimateurs comme les estimateurs spectraux.

[1] Approximating K-means-type Clustering via Semidefinite Programming, Jiming Peng, Yu Wei, 2007