

COLLOQUIUM

Salle des Conférences de l'IECN

Mardi 20 Novembre 2001 à 17H00

Pascal Massart

Professeur, Université Paris-Sud

Inégalités de concentration et Sélection de modèles

Une situation très fréquente en statistique est l'estimation d'une relation fonctionnelle $y = f(x)$ à partir de l'observation de n réalisations indépendantes mais bruitées du couple (x, y) . La fonction f est bien sûr inconnue.

Deux approches sont généralement utilisées. La première dite "paramétrique" consiste à modéliser *a priori* la fonction recherchée à l'aide d'une famille dépendant d'un nombre fini (petit) de paramètres réels. La théorie est asymptotique ($n \rightarrow \infty$). Si le modèle est inadapté cette technique est vouée à l'échec. La deuxième dite "non paramétrique" remplace l'appartenance à une famille paramétrique par une information sur la "variabilité de f " exprimée par un contrôle de sa régularité. Le gain est une réduction de l'erreur de modèle au prix d'une plus grande dispersion de l'estimateur de f .

Le but de cet exposé est de donner un aperçu de méthodes non asymptotiques offrant un bon compromis entre les deux précédentes par la sélection adaptative de modèles dans une "liste" dépendant du nombre d'observations n . L'outil mathématique principal est une inégalité de concentration de Michel Talagrand.

Ces méthodes seront illustrées par quelques exemples dans différents contextes, simulations à l'appui.