

# School Of Statistics - SOS '08



**lundi 30 juin 2008 - vendredi 4 juillet 2008**

**&lt;b&gt; IPHC - Strasbourg&lt;/b&gt;**

## **Programme scientifique**

Le thème spécifique de SOS 2008 est : **l'analyse discriminante multivariée** ou le problème de la classification statistique (signal/fond par exemple). Trois cours traitent spécifiquement de ce sujet : une introduction aux méthodes multivariées, un cours sur les réseaux de neurones et un cours sur les arbres de décision.

Auparavant, deux cours proposent un rappel des concepts fondamentaux.

<br.

L'ensemble de ces cours est complété par des sessions pratiques. Deux sessions présentent les outils RooFit et TMVA. Et deux autres sessions illustrent la combinaison des observables dans différents domaines de physique.

<it>Pour l'instant, seule une brève description de chaque session est proposée ici. Leur contenu sera détaillé lorsqu'il sera disponible.</it>

## Concepts fondamentaux

**Concepts fondamentaux 1 :**

**Trois sessions de 1 heure 30.**

**Intervenant : Imad Laktineh**

Rappel des concepts de probabilité et de variable aléatoire qui sous-tendent toutes les manipulations de données. Ce cours a pour premier objectif d'aboutir aux règles de propagation des erreurs qui seront discutées dans des cas complexes (multivariables, corrélations, systématiques). La génération Monte-Carlo des variables aléatoires selon une loi quelconque ou particulière constitue le second objectif.

**Concepts fondamentaux 2 :**

**Trois sessions de 1 heure 30.**

**Intervenant : Pierre Lutz**

Rappel des méthodes d'estimation de paramètres (estimation ponctuelle et par intervalle) et de l'erreur associée à l'estimation. Une discussion sur les cas pratiques et difficiles (par exemple la vraisemblance binnée avec une statistique MC limitée) sera appréciée. Naturellement cette première partie sera suivie d'un rappel sur les concepts fondamentaux du test d'hypothèses (qui seront utilisés par la suite pour les analyses discriminantes).

# Thèmes spécifiques

## **Analyse discriminante : des coupures séquentielles aux méthodes multivariées**

**Deux sessions de 1 heure 30.**

**Intervenant : Harrison Prosper**

Introduction générale à l'analyse discriminante (problème de la classification) et à l'estimation de leur performance. L'intervenant montrera l'intérêt de ces méthodes comparées à la technique des coupures séquentielles et dressera un panorama de tous les outils utiles. Il traitera plus spécifiquement l'une ou l'autre méthode en prenant en compte les deux cours suivants spécialisés, l'un sur les réseaux de neurones et l'autre sur les arbres de décision.

Une évocation des pièges à éviter pour les débutants (et les autres !) sera particulièrement appréciée.

## **Réseaux de neurones :**

**Une session d'environ 2 heures.**

**Intervenant : Jérôme Schwindling**

Introduction aux réseaux de neurones et à la théorie qui les rend opérationnels (apprentissage). Un objectif est de comprendre comment choisir et entraîner son réseau en fonction du problème à traiter. Les questions d'optimisation du nombre et du type de variables à utiliser, de la composition des échantillons d'apprentissage et d'estimation des performances du réseau (incluant l'erreur) prendront une place particulièrement importante.

## **Arbres de décision :**

**Une session d'environ 2 heures.**

**Intervenant : Yann Coadou**

Introduction aux arbres de décision et à leur version ``boostée''. Comme pour le cours sur les réseaux de neurones, le principe de fonctionnement de la méthode sera présenté tout autant qu'une discussion des aspects pratiques de sa mise en oeuvre.

## Sessions pratiques

**RooFit :**

**Une session d'environ 2 heures.**

**Intervenant : Wouter Verkerke**

How to use RooFit for a variety of goals (fits, Monte Carlo simulation, modelling...). A brief reminding of the methods used that have been introduced during the previous lecture will be useful.

**TMVA :**

**Deux sessions d'environ 1 heure 30.**

**Intervenant : Andreas Höcker**

Utilisation de (ROOT)TMVA pour la mise en oeuvre des méthodes discriminantes multivariées et la comparaison de leur performance.

**Combinaisons des observables en physique des saveurs :**

**Deux sessions d'environ 1 heure 30.**

**Intervenant : Jérôme Charles**

A partir des objectifs de la physique des saveurs et de celle de l'oscillation des neutrinos (matrice MNS) et de l'exemple de CKMfitter, l'intervenant montrera comment combiner pratiquement différents résultats pour un même ajustement. Les différentes méthodes mises en oeuvre seront présentées avec leurs mérites respectifs ainsi que les problèmes concrets qui apparaissent notamment pour gérer la diversité des données d'entrée, les non-linéarités, les biais potentiels et les paramètres de nuisances.

**Combinaisons d'observables en cosmologie :**

**Une session d'environ 2 heures.**

**Intervenant : Jean-Marc Virey**

A partir des objectifs de la cosmologie observationnelle, l'intervenant illustrera plusieurs cas concrets de combinaison des mesures en cosmologie observationnelle. La mise en pratique des méthodes vues dans les cours généraux sera expliquée.

Cette session propose un parallèle évident avec la présentation des méthodes de CKMfitter mais pour un domaine de physique différent qui mettra en évidence la variété des situations d'applications des outils statistiques.