

## METHODES ROBUSTES EN STATISTIQUE

### OBJECTIF DES JOURNÉES

Le but des Journées est de se consacrer pendant une semaine à l'approfondissement d'un thème bien défini, dans un cadre favorisant rencontres et discussions. Dans cette optique, le nombre de participants est limité à 45.

Les orientations principales sont :

- acquisition de notions de base
- développements les plus importants et les plus récents
- perspectives futures et applications.

Ces journées sont ouvertes à un public de statisticiens non nécessairement spécialistes du sujet traité.

Dans ce contexte, les Journées 2012 sont consacrées aux **méthodes robustes** avec pour objectif de faire le point sur la théorie et les applications par des spécialistes de divers horizons.

Bien qu'historiquement, la problématique de la robustesse en statistique remonte à des travaux lointains de mathématiciens célèbres (Bernoulli, 1777 ; Edgeworth, 1883 ; Newcomb, 1886 ; Student, 1927, etc.), l'émergence de la statistique robuste moderne ne date que des années soixante du siècle dernier avec les travaux pionniers de Tukey (1960), Huber (1964) et Hampel (1968). Depuis cette période, de nombreux modèles et méthodes ont été réexaminés sous l'angle de la robustesse.

La prise en compte de l'impact sur les méthodes statistiques de valeurs atypiques ou de toute autre structure minoritaire présente dans les données, est d'autant plus importante à l'heure actuelle que l'on dispose de bases de données de plus en plus grandes et dont la fiabilité et la qualité sont relativement inégales. Comme l'a souligné le statisticien G. Box en écrivant "All models are wrong, some are useful", il est évident que tout modèle sous-jacent n'est qu'un reflet simplifié de la réalité. Or, l'estimation des paramètres d'un modèle n'est valable que sous certaines hypothèses qui sont bien trop souvent passées sous silence dans la pratique. Une hypothèse fondamentale pour les estimateurs dits classiques suppose que tous les individus examinés ont un comportement compatible avec le modèle sous-jacent. Or, la présence dans la population de plusieurs types de comportement non identifiés par le modèle ou l'existence de valeurs aberrantes va à l'encontre de cette hypothèse. Les recherches intégrant des méthodes statistiques robustes destinées à surmonter ces difficultés sont intéressantes tant au niveau théorique qu'au niveau pratique. Ces méthodes robustes sont également essentielles dans la détection des valeurs atypiques.

Les Journées d'Etude en Statistique se proposent de faire le point sur les fondements méthodologiques et les travaux récents de la méthodologie statistique robuste et d'aborder des applications intéressantes pour le chercheur dans des domaines variés. Partant des notions fondamentales de modèle de position, d'échelle et de modèle de régression linéaire, les Journées suivantes aborderont les développements récents de la robustesse dans des modèles statistiques plus complexes : modèles d'analyse spatiale, modèle linéaire généralisé, « sparse model », « modèle » d'échantillonnage, etc.

Une avancée significative dans la recherche récente a été la généralisation des méthodes robustes au modèle de régression généralisé (GLM) où les variables dépendantes peuvent être caractérisées par des distributions continues ou discrètes quelconques. D'un point de vue théorique, cette vision généralisée partant de l'estimateur de maximum de vraisemblance, non robuste face à la présence de valeurs atypiques, a permis d'introduire une classe générale de M-estimateurs appropriés pour les modèles GLM.

La disponibilité de bases de données de plus en plus grandes, où le nombre de variables est parfois supérieur aux nombres d'observations, a également incité les chercheurs à s'intéresser à ce qu'on appelle les méthodes « sparse ». En régression, la méthode LARS (least angle regression) est

devenue très rapidement populaire dans ce contexte. Néanmoins, cette méthode n'est malheureusement pas robuste. Plusieurs alternatives robustes à cette méthode seront également exposées durant les Journées.

Un autre centre d'intérêt concerne la théorie des sondages, pour laquelle la méthode de linéarisation basée sur les fonctions d'influence est un outil puissant de l'estimation de la variance. L'utilisation de la fonction d'influence comme mesure d'influence est toutefois discutable dans la mesure où cette fonction ne prend pas en compte le plan de sondage ; d'autres mesures telles que le biais conditionnel sont plus adaptées.

Les exposés font l'objet d'un document distribué aux participants

## **LIEU**

Les Journées se déroulent au centre de vacances du CNRS à Fréjus: [Villa Clythia](http://www.caes.cnrs.fr/vacances/nos-villages/la-villa-clythia)  
<http://www.caes.cnrs.fr/vacances/nos-villages/la-villa-clythia>

## **COMITÉ D'ORGANISATION**

Jean-Jacques DROESBEKE, Université Libre de Bruxelles  
Gilbert SAPORTA, Cnam, Paris  
Christine THOMAS-AGNAN, Université Toulouse 1

## **CONFÉRENCIERS**

Eva CANTONI (Université de Genève)  
Christophe CROUX (KU Leuven)  
Catherine DEHON (Université Libre de Bruxelles)  
Gentiane HAESBROECK (Université de Liège)  
Anne RUIZ-GAZEN ( Université Toulouse 1)

## PROGRAMME

	8H30-10H00	10H30-12H00	16H00-18H00
Lundi 15 octobre	<b>J.J.Droesbeke, C.Thomas-Agnan, G.Saporta</b> Accueil Historique du sujet	<b>C.Croux</b> Mesures empiriques M-estimateurs	<b>C.Dehon</b> Régression linéaire 1: LTS, diagnostic de valeurs aberrantes, programmes R
Mardi 16 octobre	<b>A.Ruiz-Gazen</b> Modèle de position- échelle	<b>G.Haesbroeck</b> Estimateurs robustes multivariés (MCD, S) Détection des valeurs aberrantes	<b>C.Dehon &amp; E.Cantoni</b> Régression linéaire 2: M, MM, leviers, valeurs aberrantes
Mercredi 17 octobre	<b>E.Cantoni</b> Modèles linéaires généralisés 1	<b>E.Cantoni</b> Modèles linéaires généralisés 2 : réponse binaire, réponse poissonienne	
Jeudi 18 octobre	<b>A.Ruiz-Gazen</b> Echantillonnage et robustesse	<b>G.Haesbroeck</b> Classification supervisée et non- supervisée	<b>C.Croux</b> Méthodes « sparse »
Vendredi 19 octobre	<b>C.Dehon</b> Analyse en composantes principales robuste	<b>A.Ruiz-Gazen</b> Analyses robustes de données spatiales et temporelles	

## CONDITIONS D'INSCRIPTION

Ces journées sont réservées aux membres de la SFdS à jour de leur cotisation 2012. Pour adhérer : [http://www.sfds.asso.fr/36-Comment\\_Adherer](http://www.sfds.asso.fr/36-Comment_Adherer)

## FRAIS D'INSCRIPTION

avant le 31/07/2012 :

Doctorants nés après le 1/1/1987 : 260 €, Universitaires (ou assimilés) 400 €, autres 1000 €

après le 31/07/2012 :

Doctorants nés après le 1/1/1987 : 360 €, Universitaires (ou assimilés) 500 €, autres 1200 €

Les prix indiqués sont nets. Les Journées d'Etudes en Statistique de la SFdS ne sont pas assujetties à la TVA.

- Le nombre de places étant limité, les demandes d'inscriptions seront traitées selon l'ordre d'arrivée.

Le paiement peut s'effectuer par :

- chèque à l'ordre de " SFdS-Journées d'Etude en Statistique " à envoyer à l'adresse de correspondance indiquée ci-dessous

ou

- virement exclusivement sur le compte bancaire SFdS-Journées d'Etude en Statistique ouvert à la Société Générale, agence Paris-Saint-Michel, 10 rue Thénard, 75005 Paris :

RIB 30003 03085 00037296114 34 en mentionnant le nom des personnes inscrites. Attention : ce compte est distinct de celui de la SFdS

- La facture des Journées d'Etudes en Statistique de la SFdS tient lieu de convention simplifiée (organisme de formation déclaré sous le numéro 11753108575 du 16.12.1998).

## **ANNULATION**

Le remboursement des droits d'inscription sera possible sur demande écrite parvenant avant le 1 septembre 2012. Il donnera lieu à une retenue de 60 € pour frais. Aucun remboursement ne sera effectué après cette date.

## **LOGEMENT**

Le logement et les repas seront assurés sur place en pension complète, à partir du dimanche soir. Les frais d'hébergement et de repas de l'ordre de 90 euros par jour sont **à régler sur place** (chèque ou carte bancaire).

## BULLETIN DE PRÉINSCRIPTION aux JES-SFdS 2012

Votre inscription ne sera définitive que lorsque les Journées d'Etudes en Statistique de la SFdS auront reçu le paiement ou le bon de commande au plus tard trois semaines après la date de réception de ce bulletin de préinscription  
Les bons de commande doivent être libellés à l'ordre des Journées d'Etudes en Statistique de la SFdS et expédiés à l'adresse de correspondance indiquée ci-dessous.

Merci de bien vouloir remplir ce bulletin en lettres capitales

NOM..... Prénom.....

Organisme..... Fonction.....

Adresse complète de correspondance :

.....  
.....

Code Postal :..... Ville :..... Pays :.....

tél. :..... Fax:.....

E.mail :.....

Je suis universitaire ou assimilé :                      oui                       non

Je suis doctorant né après le 1/1/1987                      oui                       non   
(joindre une copie de la carte d'étudiant)

J'envisage de régler les droits d'inscription :

par chèque :                       par virement :                       par bon de commande :

DATE et SIGNATURE :

---

### CORRESPONDANCE

JOURNÉES D'ETUDE EN STATISTIQUE

Chaire de Statistique Appliquée

CNAM

292, rue Saint-Martin

75141 PARIS CÉDEX 03

Tél : 01.40.27.27.54

Fax : 01.40.27.25.48

e-mail : gilbert.saporta@cnam.fr