

# Statistiques non-paramétriques

## M2 CEE

Pr. Philippe Polomé, Université Lumière Lyon 2

2018 – 2019



# Table des matières

Motivation

Organisation

# Definitions

- ▶ En stat, le terme "non-paramétrique" a plusieurs significations :
  1. Des **stats** dites non-paramétriques
    - ▶ Qui ne s'appuient pas sur un paramètre
      - ▶ p.e. les stat d'ordre, qui s'appuient sur le rang des observations dans l'échantillon
      - ▶ En particulier, la médiane
  2. Des techniques qui ne s'appuient pas sur une distribution en particulier
    - ▶ Dites "libre de distribution" [distribution free]
      - ▶ p.e. le test t-stat ne peut être obtenu que si  $\epsilon \sim n(.,.)$
  3. Des techniques qui ne supposent pas que le modèle est fixe dans tout l'échantillon
    - ▶ Le nombre de paramètres du modèle grandit avec la taille de l'échantillon

# Applications

- ▶ 2 grands domaines
  - ▶ Régressions
    - ▶ lorsque la forme fonctionnelle de la relation n'est pas établie
  - ▶ Tests
    - ▶ lorsque les données sont des classements
- ▶ Avantages
  - ▶ Moins d'hypothèses, donc + robustes
- ▶ Inconvénient
  - ▶ Plus d'observations sont nécessaires pour atteindre la même puissance
    - ▶ C'est la monnaie de la pièce d'avoir moins d'hypothèses

# Table des matières

Motivation

Organisation

# Plan par Ch

1. Tests non-paramétriques classiques et randomisation
2. Bootstrap
3. Graphe & Régressions
  - 3.1 Histogramme
  - 3.2 Régression à kernel
  - 3.3 Semi-paramétrique à kernel
    - ▶ Type Klein-Spady ou Ichimura
    - ▶ Modèles à coefficients aléatoires
    - ▶ Modèles partiellement linéaires
  - 3.4 LOESS “locally weighted scatterplot smoothing”
    - ▶ une technique classique

# 1. Tests

- ▶ Il s'agit souvent de tests d'indépendance
  - ▶ Entre un éch. & une distribution théorique
  - ▶ Entre 2 ou + éch.
- ▶ Ces tests sont anciens
  - ▶ Correspondant à des usages assez cadrés
    - ▶ dans un tableau
  - ▶ Des implémentations modernes facilitent/généralisent les traitements
    - ▶ Permutations / randomisations

## 2. Bootstrap

- ▶ Les techniques de tests en np utilisent le bootstrap
- ▶ La randomisation peut être vue comme un bootstrap
- ▶ Le bootstrap est une technique de test fondamentalement np
  - ▶ Nous verrons une application en contexte de régression



### 3. Régressions np

- ▶ Il n'y a pas un traitement unifié de np
  - ▶ Au contraire de MC p.e.
  - ▶ En général, np basé sur une approche graphique
    - ▶ graphes sont des représentations des données
    - ▶ S'appuyant sur des approximations locales
    - ▶ Sur le fond, traçant une courbe dans un nuage de points...
- ▶ Privilégier une approche
  - ▶ opérationnelle : on peut estimer plusieurs modèles
  - ▶ moderne
  - ▶ relativement unifiée
- ▶ On suit un auteur principalement : J. Racine
  - ▶ Méthodes présentées dans un livre opérationnel assez simple
    - ▶ Nonparametric Econometrics : A Primer (2008)
    - ▶ Dispo en ligne (google)
  - ▶ Suivant un package R "np"
    - ▶ npRmpi pour la parallélisation (pas utilisé dans ce cours)
    - ▶ crs (Categorical data Regression Spline)

# Évaluation

- ▶ Examen final écrit
  - ▶ Ça n'est pas idéal
  - ▶ Mais je n'ai pas assez de recul pour faire autre chose

## Biblio principale

- ▶ Sidney Siegel “Non-parametric Statistics for the Behavioural Sciences” – nombreuses éditions
- ▶ Racine, Nonparametric Econometrics : A Primer (2008)
  - ▶ R code to replicate examples in this primer
    - ▶ [http://socserv.mcmaster.ca/racine/primer\\_code.zip](http://socserv.mcmaster.ca/racine/primer_code.zip)
- ▶ Package np & associés
  - ▶ Site <http://socserv.mcmaster.ca/racinej/Gallery/Home.html>
  - ▶ np package manual
    - ▶ <http://cran.r-project.org/web/packages/np/vignettes/np.pdf>
  - ▶ crs package manual
    - ▶ <http://cran.r-project.org/web/packages/crs/vignettes/crs.pdf>
- ▶ + les réf sur R données dans le cours de Programmation dans R