

# **RAPPORT DE STAGE D'APPLICATION EN STATISTIQUES DE 2A**

---

**STRUCTURE D'ACCUEIL : OBSERVATOIRE  
MIDI-PYRÉNÉES**

**THÈME DU STAGE : BLABLABLA**



# Introduction

## Éléments de physique et de cosmologies nécessaires à la compréhension des enjeux du sujet

Comment sommes-nous passés d'un univers sans matière à un univers structuré, via la seule gravitation, en seulement 15 milliard d'années ? Pour bien comprendre les enjeux de cette question, intéressons nous d'abord à la structure de l'univers.

La force de gravité, ou gravitation, est la force à l'origine du regroupement des particules de gaz qui forment les étoiles, du regroupement des étoiles qui forment les galaxies, et du regroupement des galaxies et de diverses matières qui forment des structures de tailles bien larges. Ces groupements de galaxies forment de longs filaments, parsemés de vide, ressemblant à une toile d'araignée. C'est ainsi qu'on se réfère couramment à la *Toile Cosmique*. En langage scientifique, nous préférons l'appellation de **Large Scale Structure** (LSS) de l'univers, qui fait donc référence à cet ensemble de structures de galaxies, sur des échelles bien plus grandes que le simple amas de galaxies.

Lorsque l'on observe la LSS de l'univers, nous observons les étoiles dont la lumière nous parvient. Cependant, la luminosité des étoiles, ainsi que leurs distances à la Terre, jouent un rôle dans notre capacité à les observer. Ainsi, pour savoir à quelle distance un objet céleste se trouve de la Terre, on ne peut pas se contenter d'observer sa luminosité.

Du fait de l'expansion de l'univers et de la dilatation de l'espace-temps induite, un décalage vers les grandes longueurs d'ondes (et donc vers le rouge pour la lumière visible) du spectre des objets lointains est observé. On parle alors de **décalage vers le rouge** (ou *redshift*). Pour une galaxie lointaine, la mesure du redshift de l'objet permet d'avoir une idée de sa distance. Pour les galaxies les plus proches, leur mouvement propre est non négligeable devant leur mouvement induit par l'expansion de l'univers et il faut donc utiliser une autre méthode de calcul des distances.

Lorsque l'on observe les différentes valeurs de redshift des objets célestes, on se rend compte que la vitesse d'expansion de l'univers est non linéaire, et plus particulièrement que cette expansion accélère. Ainsi, pour observer les objets les plus anciens, on regarde ceux qui ont le plus grand redshift. En particulier, lorsque l'on observe les objets ayant un redshift de 300, nous observons en réalité 380 000 ans après le big bang. L'espace est alors à une température de 3000K et est assez froid pour permettre aux premiers atomes de se créer. C'est le **rayonnement fossile**, ou **fond diffus cosmologique**.

Ce rayonnement correspond à la plus ancienne lumière observable depuis la Terre. Son observation nous renseigne sur l'état d'homogénéité et d'isotropie spatiale de l'univers primordial. Nous apprenons ainsi que l'univers primordial est très homogène mais présente de légères variations qui, au fil de son expansion, ont résultées en la création de la Toile Cosmique. Une meilleure connaissance de l'univers initial nous permettrait de réaliser de meilleures modélisations de son évolution au fil du temps, et donc résulterait en une meilleure connaissance de la répartition des objets célestes dans l'espace actuel.

## Contexte de l'étude

Ici, expliquer en quoi la connaissance de la répartition des lentilles gravitationnelles est reliée à la connaissance de l'homogénéité de l'univers primitif.

## Présentation des données CFHTLS-T0007

Pour ce travail, nous disposons des données issues de la 7<sup>ème</sup> et dernière version du sondage spatial *Canada-France-Hawaii Telescope Legacy Survey*. Plus précisément, nous exploiterons les données du CFHTLS-T0007 Wide, constitué de 171 pointeurs profonds dits MegaCam, chacun fournissant des mesures sur  $1 \times 1 \text{ deg}^2$ <sup>1</sup>. Cet ensemble de MegaCam produit une cartographie des objets célestes d'une taille approximative de  $155 \text{ deg}^2$  répartie sur 4 parcelles distinctes (cf. W1-4 sur [Figure 1](#)).

Les objets observés sont des galaxies occupant un cône de lumière projeté sur le ciel. Au premier ordre, ces galaxies sont divisibles en deux populations : une population d'avant plan et une population d'arrière-plan. Nous considérons les objets d'avant-plan comme ceux ayant un redshift inférieur à 0.3, et

---

1. Chaque MegaCam observe sur un champs  $1 \text{ deg} \times 1 \text{ deg}$  avec  $0.186''/\text{pixel}$  ( $19354 \times 19354$  pixels)

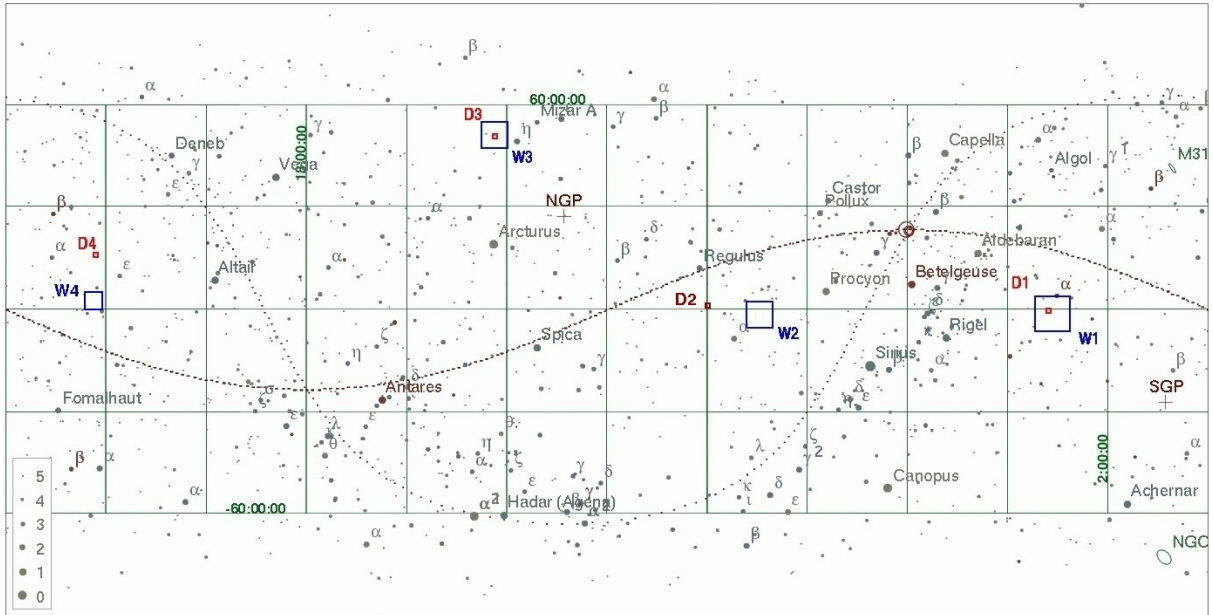


FIGURE 1 – Disposition des 4 champs larges W1-4 dans le ciel

Source: [Pat+12]

ceux d'arrière plan ayant un redshift supérieur à  $z \approx 0.5$ .

En de rares occasions, certaines galaxies d'arrière-plan sont très exactement alignées géométriquement avec des galaxies d'avant-plan dans notre ligne de visée et créent ce qu'on appelle des mirages gravitationnels. L'ensemble formé de la galaxie d'avant-plan et du mirage gravitationnel de la source d'arrière-plan est appelé **lentille gravitationnelle**. Nous disposons d'environ 130 lentilles gravitationnelles réparties sur 4 champs du ciel couvrant  $140 \text{ deg}^2$  parmi les millions d'objets. Nous considérons que notre échantillon observé est complet et pur, nous avons observé l'ensemble des lentilles gravitationnelles présentes dans le plan.

La première mission de ce stage est d'étudier la répartition spatiale des lentilles gravitationnelles. En effet

# Table des matières

---

	Page
<b>1 Étude du biais d'observation</b>	<b>4</b>
<b>A Annexe - biais d'observation</b>	<b>5</b>

# 1 Étude du biais d'observation

On commence par étudier un potentiel biais d'observation des lentilles gravitationnelles. Pour ce faire, on étudie la répartition des lentilles en fonction des valeurs de seeing et de temps d'exposition.

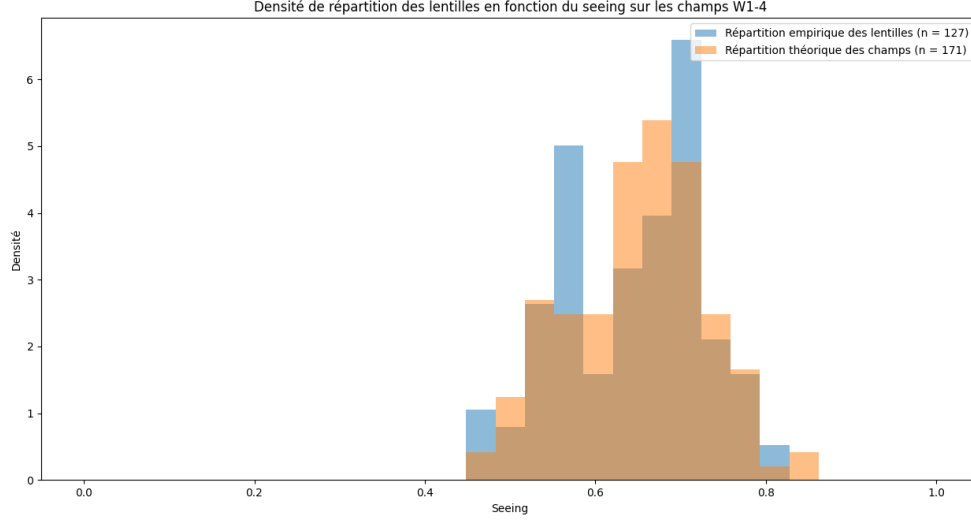


FIGURE 2 – Caption

Source: Données CFHTLS [Pat+12]

L'objectif est de voir si la répartition des lentilles en fonction du seeing suit la même loi que la répartition théorique du seeing des mesures.

Ici, comme l'on compare des distributions de variables quantitatives, l'idée première est de supposer les échantillons comme normaux (loi des grands nombres) puis de comparer leurs moyennes via des tests de student.

Cependant, ici, les analyses en qq-plot de nos échantillons (Figure 3) montre sans débat possible que ces derniers ne sont absolument pas normaux. Nous allons donc utiliser des tests non paramétriques, qui ne font pas d'hypothèse sur la forme des distributions. En particulier, nous allons utiliser le test des rangs signés de Wilcoxon [Wil45] qui permet la comparaison de deux distributions de variables continues.

Plus précisément, pour 2 échantillons  $X = (X_1, \dots, X_n)$  et  $Y = (Y_1, \dots, Y_n)$ , on calcule la différence  $D = (X_1 - Y_1, \dots, X_n - Y_n)$ . On note  $\theta$  la médiane de l'échantillon  $D$  et on effectue le test

$$H_0 : \theta = 0 \text{ vs } H_1 : \theta \neq 0$$

On ordonne alors les  $|D_i|$  et on note  $R_i$  le rang de chaque  $D_i$  dans l'échantillon ordonné. On définit l'indicatrice  $\psi_i$  qui vaut +1 si  $D_i > 0$  et -1 si  $D_i < 0$ . Si  $D_i = 0$ , on l'exclut de l'échantillon. Après avoir effectué toutes les exclusions, notre échantillon est de taille  $m$ .

On a alors la statistique de test  $T = \sum_i R_i \psi_i$ . Sous l'hypothèse nulle,  $T$  ne suit pas une loi usuelle mais une distribution spécifique d'espérance nulle et de variance  $\frac{m \times (m+1) \times (2m+1)}{6}$ .

Le test consiste alors en le rejet de  $H_0$  si  $T > T_{\text{critique}}$  avec  $T_{\text{critique}}$  disponible dans des tables de références [Low99].

Finalement, on obtient que les deux distributions ne sont pas les mêmes

## A Annexe - biais d'observation

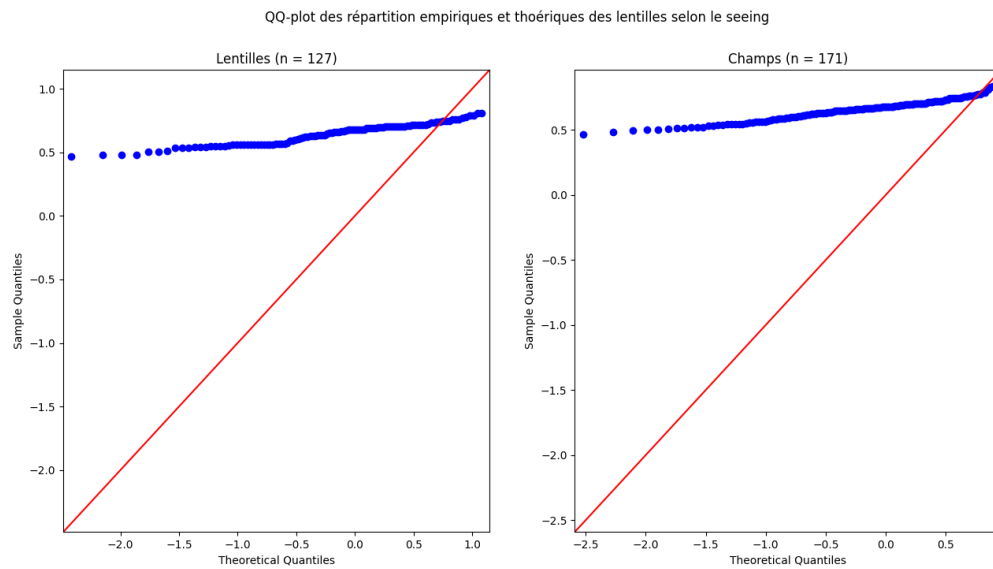


FIGURE 3 – Non normalité des échantillons en fonction du seeing

Source: Données CFHTLS [[Pat+12](#)]

## Bibliographie

- [Low99] Richard LOWRY. *Concepts Applications of Inferential Statistics - The Wilcoxon Signed-Rank Test*. 1999. URL : <http://vassarstats.net/textbook/ch12a.html>.
- [Pat+12] Hudelot PATRICK et al. *T0007 : The Final CFHTLS Release, Executive Summary*. Report. TERAPIX-CFHTLS, 2012. URL : <https://cfhtls.calet.org/T07/doc/T0007-doc.pdf>.
- [Wil45] Frank WILCOXON. « Individual Comparisons by Ranking Methods ». In : *Biometrics Bulletin* 1.6 (1945), p. 80-83. ISSN : 00994987. URL : <http://www.jstor.org/stable/3001968>.