

Analyse statistique des distributions de sortie du Kernel-Nulling pour la détection à haut contraste d'exoplanètes dans les configurations VLTI et LIFE

Vincent Foriel, David Mary, Frantz Martinache

September 24, 2025

Abstract

L'interférométrie à annulation par noyau (Kernel-Nulling) représente une approche prometteuse pour la détection directe d'exoplanètes. Cette technique génère des distributions d'intensité caractéristiques selon la présence ou l'absence d'un compagnon planétaire. L'analyse statistique de ces distributions est essentielle pour une détection robuste de planètes. Nous développons et comparons plusieurs tests statistiques pour discriminer efficacement entre les hypothèses \mathcal{H}_0 (étoile seule) et \mathcal{H}_1 (système étoile-planète). Nous analysons les performances de différentes statistiques de test afin d'évaluer la pertinence de chacun d'eux en fonction de la nature des données à disposition. Nous effectuons des simulations numériques pour trois scénarios instrumentaux : les configurations VLTI au sol et celle de LIFE dans l'espace ainsi qu'un scénario fictif consistant à placer le VLTI dans l'espace afin d'isoler les pertes de performance causées par l'atmosphère. Pour chaque scénario, nous générons des jeux de données sous les deux hypothèses \mathcal{H}_0 et \mathcal{H}_1 , en tenant compte des paramètres instrumentaux spécifiques et des niveaux de bruit. Nous évaluons les performances des tests en utilisant les courbes ROC et l'analyse des valeurs P. **Ajouter les résultats**. Cette analyse statistique, actuellement appliquée à des données simulées, ouvre la voie à une détection robuste d'exoplanètes à haut contraste utilisant l'interférométrie à annulation par noyau.

OK résumé de mes coms détaillées ci-dessous : le plus gros boulot à faire pour l'itération suivante c'est :

- en intro : biblio : recenser, comparer, contraster pour mettre en évidence l'originalité du papier
- en intro : justification de l'intérêt scientifique du papier (/exoplanètes, VLTI, LIFE) et de son scope (portée attendu des simus ?)
- Sec. 2 étude statistique plus poussée des différents régimes, types de distribution, influence des paramètres, classement et présentation des résultats pour justifier la section d'après (les tests mis en opuvre)
- Sec. 3 : La formalisation rigoureuse des tests, classement, formules analytiques
- Sec. 4 : Etude comparative des performances dans les résultats

1 Introduction

L'imagerie directe d'exoplanètes demeure l'un des défis majeurs de l'astronomie moderne, nécessitant des techniques capables de surmonter les contraintes de contraste (au-delà de 10^{-8} pour permettre la détection d'exo-Terres) et les exigences de séparation angulaire (de l'ordre de milli-arcseconde). L'interférométrie à annulation, initialement proposée par [BM79], utilise l'interférence destructive pour supprimer la source sur l'axe (lumière stellaire) tout en préservant celles hors axe (signaux planétaires), répondant ainsi aux défis de résolution angulaire et de contraste.

Le Kernel-Nulling ([MI18]) améliore cette approche en se concentrant sur la différence entre des paires de combinaisons symétriques de faisceaux qui sont robustes aux aberrations de phase du premier ordre ([MI18]). En raison de différentes sources de perturbations, la sortie de l'opération d'annulation - ainsi que l'opération de Kernel-Nulling - est une distribution statistique d'intensités. Cette distribution est influencée par divers paramètres instrumentaux tels que les erreurs de cophasage en entrée, les erreurs d'amplitude, le nombre de trames, le fond de ciel ou même d'autres sources de lumière

indésirables comme, par exemple, étoiles d'arrière-plan ([HMM⁺11], [C⁺22]). La présence d'un compagnon induit une erreur de cophasage systématique qui ne peut être approximée par une perturbation du premier ordre, conduisant à un décalage global dans les distributions de sortie du Kernel-Null. Cela résulte en des distributions statistiques distinctes pour la sortie du kernel-null selon qu'un compagnon soit présent ou non (voir Fig. 1). Dans des scénarios réalistes, le décalage induit par la présence du compagnon est petit comparé à la dispersion due aux différentes sources de perturbation, rendant les distributions \mathcal{H}_0 et \mathcal{H}_1 difficiles à distinguer. Le nombre de trames utilisé dans la simulation affecte également l'étalement et la lissité des distributions. Une analyse détaillée de l'influence de ces paramètres et des familles de distributions résultantes est fournie en Sec. 2.3. Cependant, ces distributions ne suivent pas les lois de probabilité standard, et leur forme analytique est inconnue a priori. Cela rend le problème de discrimination difficile, en particulier lorsque le contraste du compagnon est élevé et que les distributions se chevauchent significativement.

L'analyse de ces distributions nécessite donc des outils statistiques appropriés pour discriminer efficacement entre les deux hypothèses : \mathcal{H}_0 (étoile seule) et \mathcal{H}_1 (système étoile-planète). Dans ce travail, nous développons et comparons plusieurs tests statistiques pour optimiser la détection d'exoplanètes utilisant le Kernel-Nulling.

Il manque :

- Une biblio minutieuse de ce pb : mettre en évidence des différences entre des distributions dans un contexte exoplanètes, et plus généralement astro, ça existe où ? Comment les gens font ? Qu'est-ce que tu prends d'eux et qu'est-ce que tu vas apporter de nouveau ? Quelles différences entre les problématiques connexes que tu as trouvées et la tienne ? Expliquer points communs et différences.
- Justification de l'intérêt de ce papier par rapport aux instruments visés / VLTI et LIFE. Faire un topo sur VLTI et LIFE par rapport à la détection d'exoplanètes. Puis décrire comment l'approche kernel-nulling se compare aux autres approches de détection exoplanètes (pros/cons) ? Et enfin sur l'approche quel est le statut des méthodes prévues ? (en gros : on ne sait pas vraiment faire, ton papier est très important pour poser des algos. Mais attention contraster par rapport aux travaux de Hanot & co que je t'ai fait suivre par mail; et il y en a probablement d'autres.)
- plan du papier

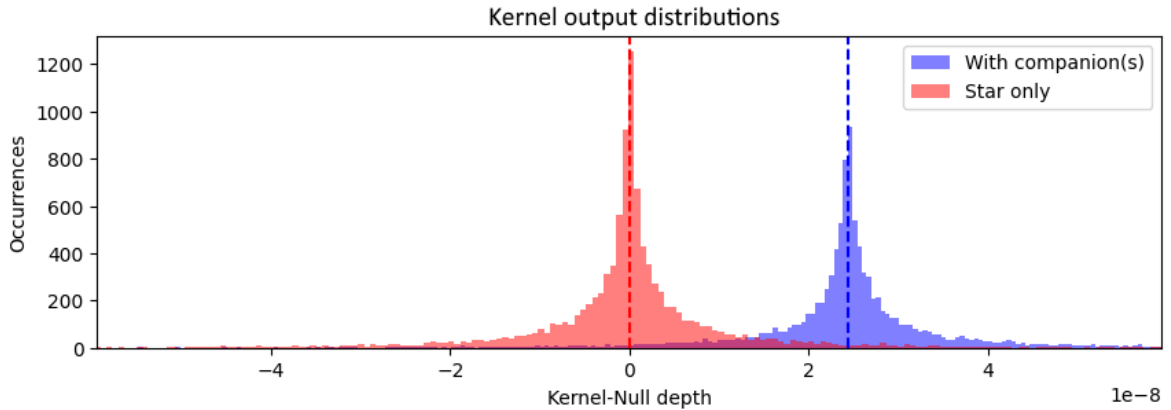


Figure 1: Exemple de distributions de profondeur d'annulation par noyau pour les hypothèses \mathcal{H}_0 (étoile seule) et \mathcal{H}_1 (avec compagnon). Ce scénario d'exemple est fortement exagéré avec un compagnon qui a un faible contraste afin d'induire un décalage significatif de la distribution. En pratique, les deux distributions sont généralement beaucoup plus proches et difficiles à distinguer. Cf texte; aussi prends l'habitude d'être ultra-spécifique, c'est un papier scientifique. Là tu es bcp trop vague : "low contrast" (c'est quoi low ?) "much closer" (=?.).

2 Méthodologie

2.1 Génération de données

Dans cette étude, nous considérons deux scénarios instrumentaux distincts pour générer des données simulées :

Paramètre	VLTI	LIFE	SVLTI
Nombre de télescopes	4	4	4
Diamètre des télescopes	8 m	2 m	8
Configuration	Irrégulière	Régulière (rectangulaire)	Irrégulière
Ligne de base maximale	130 m	600 m	130
Environnement d'exploitation	Sol	Espace	Espace
Longueur d'onde	$1.55\mu\text{m}$	$4\mu\text{m}$	$1.55\mu\text{m}$
Erreur de cophasage (RMS)	100 nm	1 nm	1 nm

Table 1: Paramètres instrumentaux pour les deux scénarios considérés dans cette étude.

Le scénario VLTI inclut une erreur de cophasage résiduelle de 100 nm RMS, représentative des conditions atmosphériques et des limitations du système de contrôle de phase au VLTI. Le scénario LIFE bénéficie de l'environnement spatial, permettant une réduction de l'erreur de cophasage à 1 nm RMS, reflétant la stabilité accrue attendue pour une mission spatiale telle que LIFE. *Il faut expliquer en détail cette table, que le lecteur ait l'impression de voir les instruments et que tes simus simulent une réalité vraiment réaliste et des manips à portée de main, ça donnera un aspect "chaud" et concret à ton papier.*

2.2 Tests statistiques et rapport de vraisemblance

Pour discriminer entre les hypothèses \mathcal{H}_0 (étoile seule) et \mathcal{H}_1 (système étoile-planète), plusieurs statistiques de test sont utilisées dans nos simulations :

- **Moyenne** : $T(u) = |\text{mean}(u)|$
- **Médiane** : $T(u) = |\text{median}(u)|$
- **Kolmogorov-Smirnov** : $T(u, v) = |D_{KS}(u, v)|$ où D_{KS} est la statistique KS entre deux échantillons
- **Flattening** : $T(u) = \sum_i |u_i - \text{median}(u)|$

Chaque test permet de quantifier la différence entre les distributions issues des deux hypothèses. Les performances sont évaluées via les courbes ROC et l'analyse des valeurs P.

Pour chaque scénario, des jeux de données sont générés sous les deux hypothèses \mathcal{H}_0 (étoile seule) et \mathcal{H}_1 (système étoile-planète), en tenant compte des paramètres instrumentaux spécifiques et des niveaux de bruit pour chaque cas. L'opération de Kernel-Nulling est supposée idéale. *Je pense qu'il faut prévoir un appendice où tu résumes le principe de kernel-nulling pour que le lecteur te suive; qu'il comprenne ce que signifie faire une opération de kernel nulling en pratique, et lister tous les défauts qui font que ça ne peut pas être idéal. Il faut aussi différencier une opération de kernel null idéale et l'absence de bruit à la fin. Bien expliquer ce que capturent tes simulations et les fluctuations qui créent les distributions montrées en Fig. 1.*

2.3 Analyse des distributions

Avant de tenter de discriminer entre les hypothèses \mathcal{H}_0 et \mathcal{H}_1 , il est essentiel d'étudier les distributions obtenues pour identifier les caractéristiques qui pourraient faciliter leur analyse. À cette fin, nous comparons les distributions simulées à différentes lois de probabilité conventionnelles, en observant notamment leur symétrie et leur forme générale.

Nous trouvons qu'aucune loi conventionnelle ne correspond parfaitement aux distributions observées. *<= Ca c'est une conclusion. Pour que cette phrase ait du poids, il faut présenter une analyse détaillée de ces distributions (comme le titre le suggère) : montrer comment les variations combinées de tous les paramètres importants mènent à des distributions différentes. Identifier et montrer tous les différents régimes de paramètres menant à des familles de distributions différentes. Lister aussi tous les paramètres qui vont entrer en compte dans les perfs de détections : contraste et position du compagnon, longueur d'onde donc, nombre de trames, puissance des erreurs de phase, D telescope,... tu peux faire une table. Une étude scientifique est vraiment une étude scientifique : on essaie d'être exhaustif dans l'analyse, et ensuite de faire une synthèse quantifiée et bien rangée qui reflète clairement tous les cas*

pour le lecteur. Le but est d'en faire un expert en lui donnant toutes les infos pour qu'ils le deviennent en lisant le papier, et puisse reproduire les résultats s'il le souhaite (c'est le propre de la démarche scientifique, éviter d'être juste déclaratif ce qui revient à proposer des arguments d'autorité; préférer systématiquement être aussi précis que possible et donner toutes les infos, ce qui revient à proposer des arguments d'expertise. Parmi les lois testées, la distribution de Cauchy semble c'est faible et pas quantifié, et vu la fig.2 ça ne colle pas comme tu le dis offrir un ajustement relativement satisfaisant (Fig. 2), bien que non parfait, en particulier sur les queues lorsque le nombre d'échantillons est élevé.

Cette observation guide le choix des tests statistiques à privilégier, en particulier ceux efficaces pour détecter des décalages dans les distributions symétriques. Justifier théoriquement et/ou empiriquement (via l'optique et le ssimus) si les distribution doivent être symétriques. Ajouter éventuellement des acquisitions de frames sur banc pour appuyer à un moment ces hypothèses. De plus, cela suggère que l'ajustement de données en minimisant \leq data fitting of what ? Pas le fit de d'une distribution empirique par une autre en tout cas... on en discutera. L'erreur quadratique moyenne (MSE) ne sera très probablement pas très efficace en raison des queues lourdes de la distribution. Au lieu de cela, nous devrions utiliser une fonction de coût dérivée de la loi de Cauchy, qui est plus robuste aux valeurs aberrantes :Non je pense que la fin de la section n'est pas bonne.

$$\text{Cost}(x, y) = \sum_i \log \left(1 + (y_i - s(x_i))^2 \right) \quad (1)$$

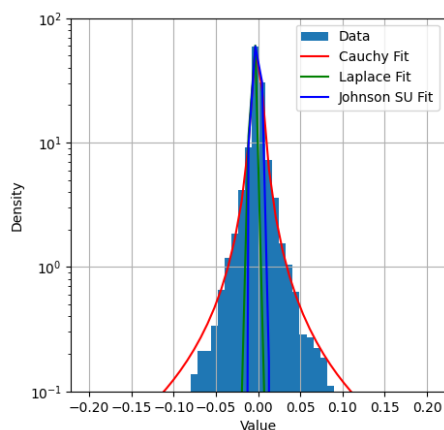


Figure 2: Ajustement des lois de probabilité conventionnelles aux distributions simulées. Le package Python Fitter a été utilisé pour effectuer l'ajustement de la plupart des lois usuelles. Cette figure montre l'ajustement des trois lois les plus pertinentes que le package a identifiées.

2.4 Tests statistiques implémentés

Nous avons développé et comparé plusieurs statistiques de test : Expliquer la démarche : vu le pb on a logiquement envie d'implémenter des tests globaux qui mesurent un shift (et rejuster que c'est juste un shift si c'est bien le cas.

- Puis dire qu'on peut aller plus loin : il est naturel de regarder des mesures plus globales entre les distributions et là tu pars sur KS etc. Ajoute Anderson-Darling dans cette famille.

- Ensuite 1) présente tes notations (par ex, qu'est-ce que x ? 2) rappelle précisément ce que tu appelles \mathcal{H}_0 et \mathcal{H}_1 en utilisant les quantités de tes notations.

- Ensuite, pour chaque test, de façon systématique : 0) Donne la ref du/des papiers scientifiques où il a été décrit (pas une ref à une toolbox python; un bouquin tu peux, mais il faut aussi les refs originales #démarche scientifique vérifiable etc) 1) explique l'intuition derrière 2) donne la stat de test (en utilisant les notations définies avant) 3) explique s'il y a une formule analytique pour décrire la stat de test sous \mathcal{H}_0 (explique aussi avant de lister les tests que l'intérêt d'une formule analytique est que ça donne accès à une p-valeur sans avoir à faire des simus de Monte Carlo, et les pbs que posent le fait de devoir faire des MC (calcul mais surtout on doit pouvoir simuler les mêmes conditions de perturbations que les données !)). Par exemple, avec le Th Centrale limite, si les x_i sont i.i.d (pas

forcément gaussiens comme dans notre cas), la stat de test (2) est gaussienne de variance connue donc pour (2) ça devrait être bon.

- Dans les cas où une distribution est disponible, vérifier par simul et montrer que la distribution empirique correspond bien à la théorique.

2.4.1 Moyenne

La statistique de test basée sur la moyenne compare la valeur absolue de la moyenne de la distribution à un seuil :

$$\left| \frac{1}{N} \sum_i x_i \right| \underset{H_0}{\overset{H_1}{\geq}} \xi \quad (2)$$

2.4.2 Médiane

Test basé sur la valeur absolue de la médiane de la distribution.

$$\begin{cases} \left| x_{\frac{N+1}{2}} \right| & \text{si } N \text{ est impair} \\ \left| \frac{x_{\frac{N}{2}} + x_{\frac{N+1}{2}}}{2} \right| & \text{si } N \text{ est pair} \end{cases} \underset{H_0}{\overset{H_1}{\geq}} \xi \quad (3)$$

2.4.3 Argmax

Cette statistique examine la position du bin avec le plus grand nombre d'occurrences dans l'histogramme des données.

Décrire formellement

2.4.4 Kolmogorov-Smirnov

Attention c'est le 2-sided KS Test comparant la distance maximale entre les fonctions de distribution cumulative des deux distributions.

Décrire formellement

2.4.5 Cramér-von Mises

Test basé sur la distance quadratique totale entre les fonctions de distribution cumulative.

Décrire formellement

2.4.6 Wilcoxon-Mann-Whitney

Test non-paramétrique pour comparer deux échantillons indépendants.

Décrire formellement

2.4.7 Aire de différence CDF

Cette statistique mesure l'aire entre les fonctions de distribution cumulative des deux distributions.

Décrire formellement

N'oublie pas le Brunner-Munzel

Il manque aussi une 3eme approche ici qui est un GLR : utiliser le modèle direct pour calculer la vraisemblance des données selon la position du compagnon; le GLR cherche alors à maximiser la position du compagnon. C'est les formules que j'avais mises au tableau dans mon bureau et la photo est sur discord. On en re-dicte aussi je sais que ça t'avait fait gamberger. C'est important parce que ce test permet la généralisation à plusieurs kernels.

D'ailleurs, il faut que tu parles de ça avant de présenter les tests : tu es en seul kernel jusqu'ici. Et il faudra ajouter une section : Generalisation des tests considérées à plusieurs kernels et/ou poses (= comment les stat de tests monokernel-obs peuvent se combiner).

2.5 Likelihood Ratio

Le test du rapport de vraisemblance (LR) consiste à comparer la probabilité d'observer les données sous chaque hypothèse :

$$\Lambda(x) = \frac{p(x; \mathcal{H}_1)}{p(x; \mathcal{H}_0)} \quad (4)$$

où $p(x; \mathcal{H}_i)$ est la densité de probabilité de la donnée x sous l'hypothèse \mathcal{H}_i .

Exemple pour une loi gaussienne : Supposons que sous chaque hypothèse, la distribution soit gaussienne de moyenne μ_i et variance σ_i^2 :

$$p(x|\mathcal{H}_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right) \quad (5)$$

Pour N observations indépendantes $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$, la vraisemblance totale sous chaque hypothèse est le produit des vraisemblances individuelles :

$$p(\mathbf{x}; \mathcal{H}_i) = \prod_{j=1}^N p(x_j; \mathcal{H}_i) \quad (6)$$

Le rapport de vraisemblance s'écrit alors :

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}; \mathcal{H}_1)}{p(\mathbf{x}; \mathcal{H}_0)} \quad (7)$$

$$= \prod_{j=1}^N \frac{p(x_j; \mathcal{H}_1)}{p(x_j; \mathcal{H}_0)} \quad (8)$$

$$= \left(\frac{\sigma_0}{\sigma_1}\right)^N \exp\left(\sum_{j=1}^N \left[\frac{(x_j - \mu_0)^2}{2\sigma_0^2} - \frac{(x_j - \mu_1)^2}{2\sigma_1^2}\right]\right) \quad (9)$$

En pratique, pour des raisons de stabilité numérique, on utilise le logarithme du rapport de vraisemblance (en négligeant les constantes qui n'ont pas d'impact sur la décision) :

$$\log \Lambda(\mathbf{x}) \propto \sum_{j=1}^N \left[\frac{(x_j - \mu_0)^2}{2\sigma_0^2} - \frac{(x_j - \mu_1)^2}{2\sigma_1^2}\right] \quad (10)$$

Ce principe est généralisé à d'autres lois (Cauchy, Laplace, etc.) selon la forme des distributions observées. Les paramètres sont estimés par ajustement sur les données simulées (voir section précédente).

Ainsi pour Laplace, dont la densité de probabilité est :

$$p(x|\mathcal{H}_i) = \frac{1}{2b_i} \exp\left(-\frac{|x - \mu_i|}{b_i}\right) \quad (11)$$

le rapport de vraisemblance devient :

$$\log(\Lambda(\mathbf{x})) \propto \sum_{i=1}^n \frac{|x_i - \mu_1|}{b_1} - \frac{|x_i - \mu_0|}{b_0} \quad (12)$$

Pour Cauchy, dont la densité de probabilité est :

$$p(x|\mathcal{H}_i) = \frac{1}{\pi\gamma_i \left[1 + \left(\frac{x - x_{0,i}}{\gamma_i}\right)^2\right]} \quad (13)$$

le rapport de vraisemblance devient :

$$\log(\Lambda(\mathbf{x})) \propto \sum_{i=1}^n \log\left(1 + \left(\frac{x_i - x_{0,1}}{\gamma_0}\right)^2\right) - \log\left(1 + \left(\frac{x_i - x_{0,1}}{\gamma_1}\right)^2\right) \quad (14)$$

3 Résultats

A la fin en bilan fais une liste des pros et cons de chaque méthode. Il faut que tu déploies dans ce papier une analyse exhaustive et convaincante de la nature des données et des régimes qu'on peut y trouver, des types de distributions suivant les régimes, et des types de tests qu'on peut utiliser. A la fin on se dira que tu as plié le problème, le spécialiste de la détection sur des données kernel-nulling c'est toi !

3.1 Courbes ROC

Les courbes ROC (Receiver Operating Characteristic) permettent de comparer l'efficacité \leq power de différentes statistiques de test en représentant la proportion de détections vraies en fonction de la probabilité de fausse alarme.

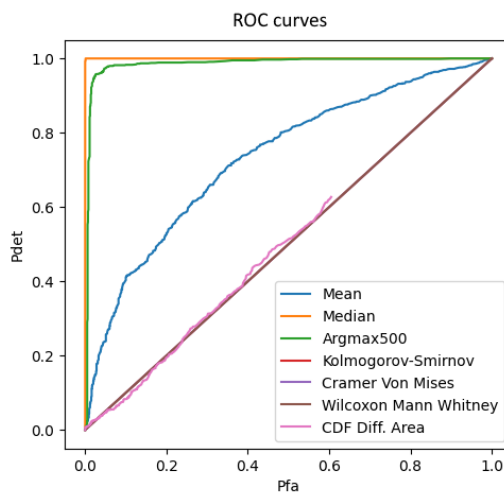


Figure 3: Courbes ROC pour différentes statistiques de test.

Analyse des résultats Ajouter la courbe de Neyman-Pearson pour la comparaison

3.2 Analyse des valeurs P

Attention Fig.4 c'est faux : ce ne sont pas les p-valeurs car les p-valeurs ne dépendent pas d'un seuil, seulement des données.

Les valeurs P fournissent une mesure de confiance pour rejeter l'hypothèse nulle. Une valeur P inférieure à 0,05 est généralement considérée comme significative. Non ça ça dépend des applications. Donne la définition précise de la p-valeur tout de suite, comme ça on sait de quoi on parle.

Pour l'approche Neyman-Pearson (ou Likelihood-Ratio), il faut en faire une section à part dans la section d'avant en explicitant le Likelihood-Ratio et les paramètres dont il dépend sous les 2 hypothèses, puis en tirer la stat de test que ça donne en prenant le log etc . Expliquer aussi l'intérêt du NP..

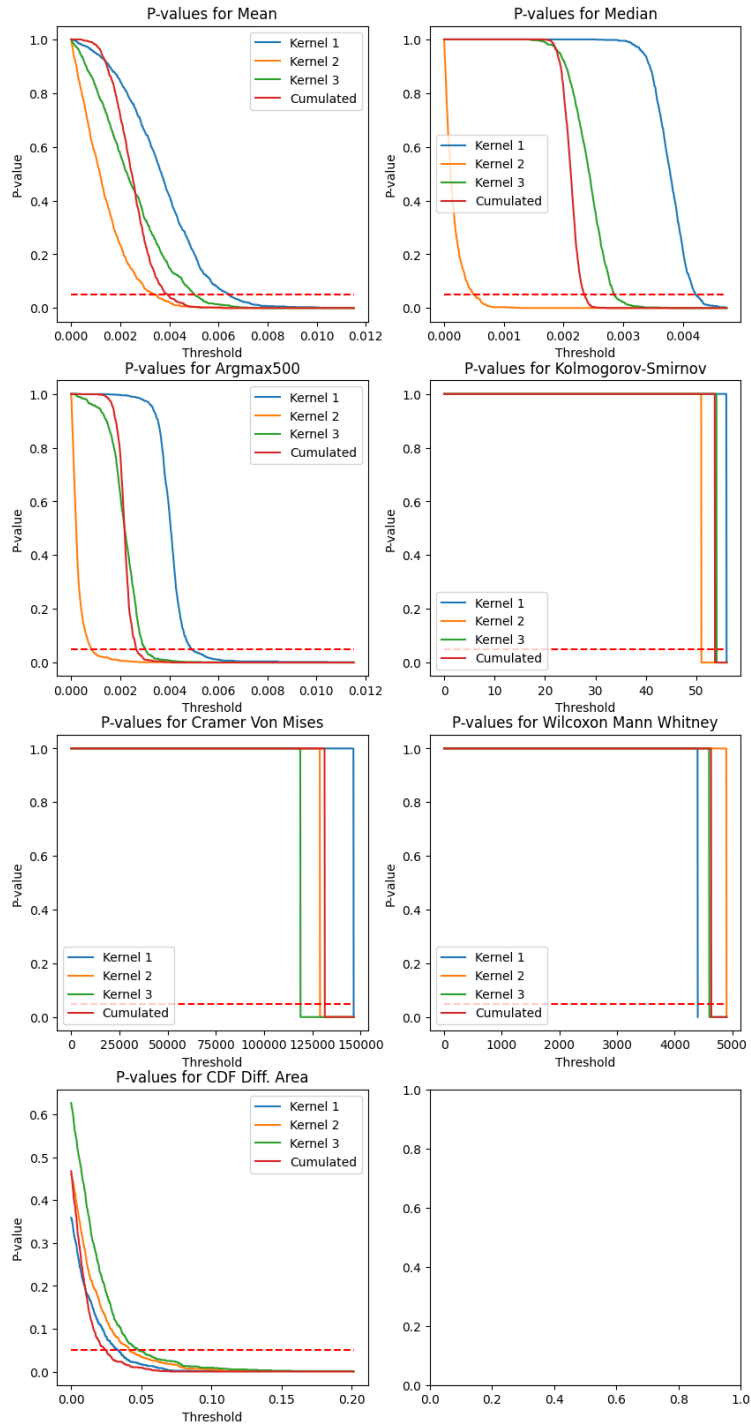


Figure 4: Évolution des valeurs P en fonction du seuil pour différentes statistiques de test. La ligne pointillée rouge indique le seuil de significativité à 0,05. **Plot à refaire pour se focaliser sur un seul Kernel (+ correction des bugs)** Oui. On peut par exemple montrer la calibration théorique p-valeur(stat de test - pas seuil !) et comparer à la calibration empirique pour chaque test

Analyse des résultats

4 Discussion

4.1 Performance comparative des tests

ToDo

4.2 Sensibilité au bruit

ToDo

5 Conclusions

ToDo

References

- [BM79] Ronald N. Bracewell and Robert H. MacPhie. Searching for nonsolar planets. *Icarus*, 38(1):136–147, April 1979.
- [C⁺22] N. Cvetojevic et al. 3-beam self-calibrated kernel nulling photonic interferometer, 2022. Preprint.
- [HMM⁺11] C. Hanot, B. Mennesson, S. Martin, K. Liewer, F. Loya, D. Mawet, P. Riaud, O. Absil, and E. Serabyn. Improving interferometric null depth measurements using statistical distributions: Theory and first results with the palomar fiber nuller. *The Astrophysical Journal*, 729(2):110, 2011.
- [MI18] Frantz Martinache and Michael J. Ireland. Kernel-nulling for a robust direct interferometric detection of extrasolar planets. *Astronomy & Astrophysics*, 619:A87, 2018.