

# QUELQUES TESTS DE DÉTECTION S'ADAPTANT À LA DISTRIBUTION DU BRUIT DE FOND EN ASTRONOMIE

David Mary<sup>1</sup>, Étienne Roquain<sup>2</sup>, Marie Perrot-Dockes<sup>2</sup>, Sophia Sulis<sup>3</sup> & Sébastien Bourguignon<sup>4</sup>

<sup>1</sup> *Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Boulevard de l'Observatoire, CS 34229, 06304, Nice cedex 4, France; david.mary@unice.fr*

<sup>2</sup> *Laboratoire de Probabilités, Statistique et Modélisation (LPSM, UMR 8001), Faculté des Sciences et Ingénierie, Tour 15/16, Étage 2, BC 158, 4, place Jussieu, 75252, PARIS cedex 05; etienne.roquain@upmc.fr, marie.perrocks@gmail.com*

<sup>3</sup> *Aix Marseille Université, CNRS, CNES, LAM, Marseille, France; Sophia.Sulis@lam.fr*

<sup>4</sup> *Laboratoire des Sciences du Numérique de Nantes (LS2N, UMR 6004), École Centrale de Nantes, 1 rue de la Noë, BP 92101, 44321 Nantes Cedex 3; Sebastien.Bourguignon@ec-nantes.fr*

**Résumé.** Dans cette note, nous considérons le problème de construction de tests de détection pour certains types de données issues de l'astrophysique. Une caractéristique commune des données considérées est que la distribution du bruit de fond est inconnue, ce qui invalide l'utilisation de nombreux tests classiques. Nous présentons des résultats récents proposant des solutions à ce problème pour deux applications spécifiques, la détection d'exoplanètes et celle de galaxies.

**Mots-clés.** détection, tests multiples, exoplanètes, galaxies.

**Abstract.**

In this note, we consider the problem of building detection tests for some types of astrophysical data. The distribution of the background noise is unknown in the considered cases, preventing from using most classical procedures. We present works that address this problem for two specific applications : the detection of exoplanets and of galaxies.

**Keywords.** detection, multiple testing, exoplanets, galaxies.

## 1 Problématique et contexte

La détection de sources est un enjeu majeur dans plusieurs domaines de l'astrophysique. Le contexte moderne d'instruments toujours plus complexes produisant des données toujours plus volumineuses conduit souvent à réaliser un nombre gigantesque de tests simultanément. La thématique des tests multiples permet de prendre en compte cette multiplicité de manière appropriée. Si ce domaine de recherche possède des origines très

anciennes en statistique, les deux dernières décennies y ont vu une explosion de travaux théoriques et appliqués : parmi les procédures qui ont vu le jour, on peut citer notamment la procédure de Benjamini-Hochberg [1] qui contrôle le *false discovery rate* (FDR), ainsi que les différentes versions du *higher criticism* (HC) [2] et des tests de Berk-Jones (BJ) [3] pour le cas de signaux rares et faibles [2].

En astrophysique, ce genre de procédures reste cependant assez peu utilisé. Une des raisons est que la distribution des données du bruit de fond (c'est-à-dire lorsque qu'il n'y a pas de signal astrophysique) est très souvent inconnue. Ceci pousse l'utilisateur à se tourner vers des procédures plus *ad hoc*, ce qui peut s'avérer dangereux. Un exemple récent est le cas de la détection d'une exoplanète près de l'étoile  $\alpha$  Centauri Bb [4]. La détection de cette exoplanète, annoncée en 2012 sur la base d'une  $P$ -valeur évaluée à 0.02% [4], a été fortement remise en cause depuis [5, 6]. En effet, ces dernières analyses ont mis en lumière des effets mal pris en compte, en particulier le signal parasite émis par l'étoile elle-même (le "bruit stellaire"). Ces effets, très difficiles à contrôler, impactent fortement les taux d'erreur prédits par les tests qui les ignorent.

L'objet de cette note est de présenter des procédures de détection avec un risque correctement contrôlé même lorsque la distribution des statistiques de test sous la distribution nulle est mal connue. Nous considérons deux cas concrets : la détection d'exoplanètes par vélocimétrie radiale [7] (Section 2), et la détection de galaxies dans des images multi-longueurs d'onde de l'instrument MUSE [8, 9] (Section 3).

## 2 Détection d'exoplanètes par vitesses radiales

La présence d'une planète orbitant autour d'une étoile induit un mouvement de l'étoile autour du barycentre de masse du système étoile-planète. Ce mouvement module de façon quasi-périodique la vitesse radiale de l'étoile par rapport à un observateur terrestre et s'imprime par effet Doppler sur la lumière de l'étoile. En mesurant le décalage Doppler des raies du spectre stellaire en fonction du temps, on déduit ainsi la vitesse radiale de l'étoile. Les données se présentent donc sous la forme d'une série temporelle obtenue durant une fenêtre temporelle d'observation.

Le but est de tester, dans une telle série de vitesses radiales,  $H_0$  : la moyenne est nulle (il n'y a pas d'exoplanète) contre  $H_1$  : la moyenne est un signal quasi-périodique (il y a une ou plusieurs exoplanètes). La série temporelle est supposée stationnaire, mais sa matrice de covariance  $\Sigma$  est inconnue en raison du bruit stellaire. Pour s'adapter à l'alternative, une approche classique en échantillonnage régulier consiste à construire le périodogramme des données (module carré de la transformée de Fourier discrète de la série), puis à rechercher si ses composantes indiquent de façon significative la présence de signaux périodiques. Cependant, la distribution de ce périodogramme est inconnue sous  $H_0$  car  $\Sigma$  est inconnue.

Pour contourner cette difficulté nous proposons ici une approche exogène, rendue pos-

sible par la disponibilité d'un certain nombre ( $L$ ) de séries simulées sous  $H_0$ . En effet, il existe à l'heure actuelle des codes astrophysiques permettant de simuler de façon réaliste les vitesses radiales qui seraient observées pour un type d'étoile donné. Ceci permet, au moins sur une certaine plage de fréquences, de générer des séries exogènes sous  $H_0$ . Celles-ci sont cependant en nombre très limité (quelques unités à une dizaine) en raison du fort coût de calcul que ces simulations nécessitent. A l'aide de ces séries, nous proposons de construire un périodogramme de référence, utilisé pour normaliser le périodogramme des données.

Notons  $Z_1, \dots, Z_N$  un sous-ensemble de composantes du périodogramme ainsi standardisé. Si l'échantillonnage est régulier, alors les distributions considérées possèdent des expressions analytiques [10]. Ainsi, sous certaines conditions, on peut montrer que  $Z_1, \dots, Z_N$  sont asymptotiquement indépendants avec  $Z_j \sim \mathcal{F}(2, 2L)$  (loi de Fisher de paramètres 2 et  $2L$ ) sous l'hypothèse nulle et  $Z_j \sim \mathcal{F}_{\lambda_j}(2, 2L)$  (loi de Fisher décentrée avec un certain de paramètre de décentrage  $\lambda_j$  et de paramètres 2 et  $2L$ ) sous l'alternative. Soulignons que sous  $H_0$ , la distribution du périodogramme standardisé est indépendante du paramètre de nuisance  $\Sigma$ . Cette propriété, complétée par celle de l'indépendance asymptotique, permet de construire des tests globaux de niveau  $\alpha$  à partir des  $Z_j$ . Par exemple, une façon naturelle et explicite d'agréger ces tests est simplement le test du maximum. Dans le cas considéré, celui-ci rejette l'hypothèse nulle si  $\max(Z_j, 1 \leq j \leq N)$  dépasse le seuil  $c_\alpha$ , calibré de sorte que

$$1 - \left(1 - \left(\frac{L}{c_\alpha + L}\right)^L\right)^N = \alpha.$$

Des résultats similaires peuvent être obtenus pour de nombreux autres tests comme celui de Fisher [11], ses variantes [12, 13, 14] et ceci permet également d'utiliser des tests de détection de type HC et BJ.

Sur des simulations, on voit que l'approximation asymptotique est bonne quand la longueur de la série est suffisamment grande devant la durée caractéristique de corrélation du signal (typiquement un à deux ordres de grandeur supérieur), ce qui confirme que le test conduit est bien de niveau  $\alpha$ . De plus, les modèles analytiques obtenus permettent d'étudier la puissance asymptotique de cette procédure de test. Là aussi, les simulations montrent que ces expressions sont valables pour des valeurs de  $N$  modérées. Ces résultats permettent de faire des études de détectabilité pour des planètes ou des instruments avec des caractéristiques données (cf Figure 1 pour un exemple avec le test du maximum).

Par ailleurs, dans le cas de l'échantillonnage irrégulier, l'hypothèse d'indépendance entre les  $Z_j$  ne peut être tenue et nous avons proposé dans [15] des techniques de *bootstrap* pour approcher la distribution de la statistique de test sous  $H_0$ .

### 3 Détection de galaxies dans les données MUSE

Le spectrographe intégral de champ MUSE installé sur un des télescopes de 8m au Very Large Telescope (Chili) permet d’obtenir des images multi-longueurs d’onde (typiquement  $300 \times 300$  images dans 3600 canaux optiques). On cherche à détecter dans ce “cube” de données des galaxies très lointaines, qui se manifestent comme une surbrillance locale du flux (une raie en émission) dans une poignée de voxels. On connaît à peu près la forme de ces raies mais ni leurs hauteurs, ni leur nombre (quelques dizaines à quelques centaines typiquement) ni leurs positions dans le cube. Par ailleurs, la hauteur de certaines raies peut être beaucoup plus faible que le niveau du bruit de fond et que celui d’autres sources brillantes et étendues (dûes à d’autres étoiles et galaxies ou à des artefacts instrumentaux; on regroupe ces sources sous le terme de signaux de nuisance).

Dans ce cadre, nous avons considéré une approche de détection en deux temps : les signaux de nuisance sont d’abord supprimés et l’étape de détection des galaxies se fait ensuite, dans le résidu. Pour s’adapter à l’alternative, l’approche considère les maxima locaux du cube de données résiduel, après suppression des nuisances et divers filtrages. En raison de ces prétraitements, la distribution sous  $H_0$  des maxima locaux n’est pas connue. Le problème considéré ici est plus complexe que le précédent puisqu’il y a plusieurs hypothèses nulles, chacune liée à un maximum local [16]. Si nous notons  $x, y, z$  la position d’un maximum local, on teste  $H_{0,x,y,z}$  : il n’y a pas de raie d’émission à la position  $(x, y, z)$ , contre  $H_{1,x,y,z}$  : il y en a une à cette position. Le critère d’erreur considéré est celui du FDR, moyenne du taux de fausses découvertes parmi les positions déclarées comme correspondant à une galaxie.

L’approche que nous proposons pour ce problème est endogène. Nous montrons par simulations que sous certaines hypothèses, la distribution sous  $H_0$  peut être estimée à partir des données elles-mêmes et que la procédure de détection globale (incluant la suppression des nuisances et les divers filtrages) contrôle le FDR. Nous appliquons cette approche à un cube de données issu de l’instrument MUSE et comparons avec les résultats du télescope spatial Hubble [9].

### 4 Perspectives

Dans la communauté statistique mathématique, la validation de procédures de test apprenant automatiquement la distribution nulle remonte aux procédures de test par permutation ou par randomisation. Dans le cadre de tests multiples et du FDR, certaines justifications ont été apportées récemment, voir par exemple [17, 18, 19]. Dans certaines des études considérées ici, la validation des approches est faite par des simulations numériques, qui nécessitent un choix particulier (et donc nécessairement un peu arbitraire) du jeu de paramètres. Une direction de recherche de ces travaux est d’obtenir des garanties théoriques en simplifiant éventuellement les procédures et/ou les modèles astrophysiques tout en s’inspirant des résultats théoriques récents.

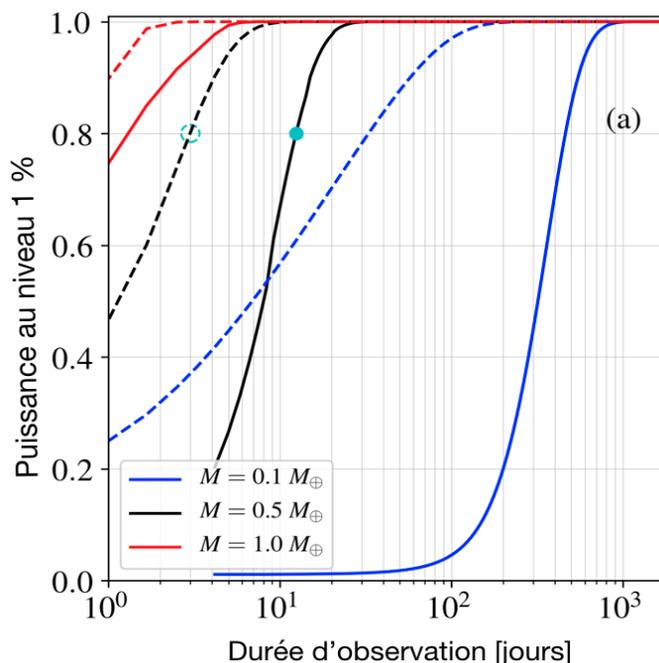


Figure 1: Puissance du test de détection au niveau 1% en fonction du temps pour une planète orbitant circulairement une étoile de type solaire en 17,5 h. La masse de l'exoplanète est de 0,1 (bleu), 0,5 (noir) et 1 (rouge) masse terrestre. Le pas d'échantillonnage des données est de 2h et  $L = 20$ . Les courbes pleines correspondent au cas où le bruit généré par l'étoile est corrélé (cas réaliste) et les courbes en pointillés au cas où il est blanc. L'écart-type du bruit est de  $49 \text{ cm.s}^{-1}$  et l'amplitude du signal planétaire de quelques centimètres par seconde. La détection est plus difficile dans le cas d'un bruit coloré (courbes pleines) car la période de la planète correspond à une zone de fréquences où le bruit de fond créé par l'étoile est très énergétique. Dans cette configuration, la détection d'une exoplanète d'une demi masse terrestre serait possible avec une probabilité de 80% en 12,4 jours; pour un bruit de même puissance mais blanc, cette durée tomberait à 3 jours (disque et cercle bleus).

## References

- [1] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300, 1995.
- [2] D. Donoho and J. Jin. Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Stat.*, 2004.
- [3] V. Gontcharuk et al. The intermediates take it all: Asymptotics of higher criticism statistics and a powerful alternative based on equal local levels. *Biom. J.*, 57(1):159–180, 2014.

- [4] X. Dumusque et al. An Earth-mass planet orbiting  $\alpha$  Centauri B. *Nature*, 491:207–211, 2012.
- [5] A. Hatzes. The Radial Velocity Detection of Earth-mass Planets in the Presence of Activity Noise: The Case of  $\alpha$  Centauri Bb. *ApJ*, 770:133, 2013.
- [6] V. Rajpaul et al. Ghost in the time series: no planet for Alpha Cen B. *MNRAS*, 456:L6–L10, 2016.
- [7] Sophia Sulis. *Statistical methods using hydrodynamic simulations of stellar atmospheres for detecting exoplanets in radial velocity data*. Theses, Université Côte d’Azur, October 2017.
- [8] R. et al. Bacon. The muse hubble ultra deep field survey - i. survey description, data reduction, and source detection. *A&A*, 608:A1, 2017.
- [9] D. Mary, R. Bacon, S. Conseil, L. Piqueras, and A. Schutz. Origin: Blind detection of faint emission line galaxies in muse datacubes. *Astronomy Astrophysics*, Jan 2020.
- [10] S. Sulis, D. Mary, and L. Bigot. A study of periodograms standardized using training datasets and application to exoplanet detection. *IEEE Transactions on Signal Processing*, 65(8):2136–2150, April 2017.
- [11] R.A. Fisher. Tests of Significance in Harmonic Analysis. *Proc. R. Soc. London, Ser. A*, 125:54–59, 1929.
- [12] S.T. Chiu. Detecting periodic components in a white gaussian time series. *J. R. Stat. Soc. Series B*, 51(2):249–259, 1989.
- [13] M. Shimshoni. On fisher’s test of significance in harmonic analysis. *Geophys. J. R. Astronom. Soc.*, pages 373–377, 1971.
- [14] E. Bölviken. New tests of significance in periodogram analysis. *Scandinavian J. Stat.*, 10(1):1–9, 1983.
- [15] S. Sulis, D. Mary, and Lionel Bigot. A bootstrap method for sinusoid detection in colored noise and uneven sampling. application to exoplanet detection. In *EUSIPCO 2017, Kos, Greece, August 28 - September 2, 2017*, pages 1095–1099. IEEE, 2017.
- [16] Dan Cheng and Armin Schwartzman. Multiple testing of local maxima for detection of peaks in random fields. *Ann. Statist.*, 45(2):529–556, 04 2017.
- [17] Rina Foygel Barber and Emmanuel J. Candès. Controlling the false discovery rate via knockoffs. *Ann. Statist.*, 43(5):2055–2085, 2015.
- [18] Ery Arias-Castro and Shiyun Chen. Distribution-free multiple testing. *Electron. J. Stat.*, 11(1):1983–2001, 2017.
- [19] Etienne Roquain and Nicolas Verzelen. On using empirical null distributions in Benjamini-Hochberg procedure. *arXiv e-prints*, page arXiv:1912.03109, Dec 2019.