

Détection d'exoplanètes par imagerie directe, une approche Bayésienne

Isabelle SMITH, André FERRARI

Laboratoire Fizeau, UNS/OCA/CNRS,
Parc Valrose, 06108 Nice cedex 2, France

isabelle.smith@unice.fr, ferrari@unice.fr

Résumé – Nous proposons et appliquons un modèle Bayésien hiérarchique pour la détection d'exoplanètes par imagerie directe. Ce modèle prend en compte l'utilisation d'images à plusieurs longueurs d'ondes, la présence de speckles (quasi-)statiques et une corrélation spatiale et inter-spectrale des bruits. Deux familles de tests Bayésiens sont étudiées : la probabilité a posteriori du rapport de vraisemblance (PLR) et le facteur de Bayes fractionnaire (FBF). Ces tests sont implémentés en utilisant une méthode d'échantillonnage. Des simulations différentes sur différents types de données mettent en évidence les bonnes performances de ces tests par rapport à des tests plus classiques.

Abstract – We propose and apply a hierarchical Bayesian model for the exoplanet detection task using direct imaging. This model accounts for multispectral images, (quasi-)static speckles and a spatial and interspectral noise correlation. Two Bayesian families are studied: the Posterior probability of the Likelihood Ratio (PLR) and the Fractional Bayes Factor (FBF). These tests are implemented using a sampling method. Different simulations on different kinds of data show the good performances of these tests when compared to more classical tests.

1 Détection d'exoplanètes par imagerie directe

La recherche de planètes en orbite autour d'étoiles autres que le Soleil –exoplanètes– est d'un grand intérêt scientifique, en témoigne le grand nombre de projets instrumentaux dévoués à cette tâche [1]. Dans le proche infra-rouge, le contraste entre la lumière diffractée par l'étoile et la planète est de 10^7 pour une planète géante et 10^{10} pour une planète du type de la Terre. De plus, compte tenu de la distance à l'observateur, la séparation angulaire étoile-planète peut descendre jusqu'à 10^{-4} deg. Ainsi, la détection d'exoplanètes par imagerie directe revient à détecter une source très faible dans un halo très brillant et bruité, ce qui explique que la première détection par cette technique ne date que de 2005 [1].

De grands projets instrumentaux en cours de développement seront désormais aptes à relever le défi, comme les futurs instruments au sol SPHERE (Spectro-Polarimetric High-Contrast Exoplanet REsearch) au VLT (Very Large Telescope) et GPI (Gemini Planet Imager) du télescope Gemini. Ces instruments reposent sur une optique adaptative (OA) "extrême" et un coronographe qui réduit la lumière diffractée de l'étoile au niveau de la planète. Une image simulée est présentée dans la Fig. 1. Les simulations sont effectuées par le software package SPHERE [2] qui simule finement la propagation lumineuse de la source au récepteur.

Dans une séquence d'images, la rotation apparente de la planète due à la monture alt-azimutale du VLT est connue, à la position initiale de la planète près. Ce mouvement pourra être

utilisé pour discriminer la planète des speckles quasi-statiques dus aux aberrations optiques du système [3]. Une analyse de ce type de méthodes, utilisant en plus une modélisation statistique détaillée, a été proposée dans [4]. Une solution pour améliorer les performances de la détection / estimation utilise le fait que les aberrations varient peu pour des longueurs d'ondes proches. L'acquisition des images dans VLT SPHERE se fera ainsi de façon simultanée dans deux longueurs d'onde. La source peut alors être détectée par un traitement différentiel [3], mais le problème des aberrations différentielles reste posé. Cette communication apporte une solution à ce problème en utilisant un formalisme Bayésien.

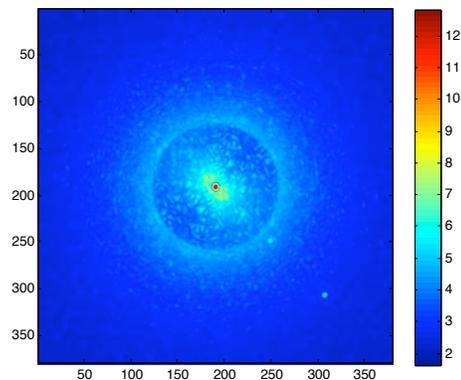


FIG. 1 – Image obtenue à partir d'une simulation end-to-end du futur instrument VLT-SPHERE. Quatre planètes sont simulées le long de la diagonale inférieure droite avec un contraste particulièrement bas de 10^4 .

2 Modèle Bayésien hiérarchique

Les données sont constituées de K ensembles de L images où chaque image est représentée par un vecteur $M \times 1$ noté $\mathbf{i}_\ell(k)$, $k = 1 \dots K$ et $\ell = 1 \dots L$. M est le nombre de pixels par image, L le nombre de canaux spectraux et K le nombre d'expositions. Afin d'améliorer les performances de détection, il est important de prendre en compte à l'instant k à la fois la corrélation spatiale (sur une image) et la corrélation interspectrale (entre deux images). Pour cela nous proposons de modéliser le vecteur $\mathbf{x}_k^t = (\mathbf{i}_1(k)^t, \dots, \mathbf{i}_L(k)^t)$. Une extension naturelle du modèle utilisé dans le cas d'une seule longueur d'onde dans [4] (dans un contexte fréquentiste) au cas multi-canaux consiste à supposer que, conditionnellement, les \mathbf{x}_k sont indépendants distribués selon

$$\mathbf{x}_k | \boldsymbol{\mu}, \Sigma, \boldsymbol{\eta}, A_k \sim \mathcal{N}_{LM}(A_k \boldsymbol{\eta} + \boldsymbol{\mu}, \Sigma) \quad (1)$$

- $\boldsymbol{\mu}$ désigne le fond *inconnu*. Dans le cas présent, il provient à la fois des résidus atmosphériques de l'étoile coronographiée et des speckles statiques liés aux aberrations optiques.
- le vecteur $\boldsymbol{\eta}$ de taille $L \times 1$ contient les intensités *inconnues* des sources dans chaque canal.

$$A_k = \begin{pmatrix} \mathbf{p}_1(k) & \mathbf{0}_{M \times 1} & \dots & \mathbf{0}_{M \times 1} \\ \mathbf{0}_{M \times 1} & \mathbf{p}_2(k) & \dots & \mathbf{0}_{M \times 1} \\ \dots & \dots & \dots & \dots \\ \mathbf{0}_{M \times 1} & \mathbf{0}_{M \times 1} & \dots & \mathbf{p}_L(k) \end{pmatrix}$$

où $\mathbf{p}_\ell(k)$ est un vecteur contenant la réponse instrumentale supposée *connue* de la source à l'instant k dans le canal ℓ . Dans le contexte astrophysique, on peut effectivement supposer que la forme du profil est connue (par calibration) et que le mouvement de la source est connu (mouvement de rotation de champ déterministe) à la position initiale près. Nous supposons afin de simplifier la présentation que la position à l'instant $k = 1$ de l'éventuelle source est connue.

- Σ est la matrice de covariance $LM \times LM$ intra et interspectrale *inconnue*.

Les intensités de la source dans les différentes bandes étant positives, plus ou moins corrélées, et surtout pouvant s'étendre sur une large plage de magnitudes, il est naturel de supposer que la distribution de $(\ln \eta_1, \dots, \ln \eta_L)^t$ est conjointement Gaussienne. La distribution du vecteur $\boldsymbol{\eta}$ est donc Log-Normale :

$$\boldsymbol{\eta} | \mathbf{m}, B \sim \log \mathcal{N}_L(\mathbf{m}, B), \quad \mathbf{m} = \mathbb{E}[\ln \boldsymbol{\eta}], \quad B = \text{cov}[\ln \boldsymbol{\eta}] \quad (2)$$

Les paramètres de nuisance de la vraisemblance et de l'a priori sont en général inconnus ou incertains. Ce problème peut être traité par l'hypothèse que ces paramètres sont aléatoires et décrits par une distribution donnée, ce qui permet de les marginaliser et de rendre les inférences sur le paramètre d'intérêt $\boldsymbol{\eta}$ plus robustes [5]. L'intégration est en général trop longue numériquement, mais elle peut-être effectuée analytiquement en décrivant les paramètres de nuisance par des distributions conjuguées.

La vraisemblance (1) et l'a priori (2) sont reliés à la distribution normale : la première avec une moyenne variable et

la seconde à travers la fonction Log. La distribution normale-inverse Wishart qui est la conjuguée pour la moyenne et la covariance de la distribution normale, est également conjuguée pour les paramètres de nuisance de (1) et (2). Nous supposons donc :

$$\begin{aligned} \boldsymbol{\mu} | \Sigma, \boldsymbol{\mu}_0, \lambda_0 &\sim \mathcal{N}_{LM}(\boldsymbol{\mu}_0, \lambda_0 \Sigma), \quad \Sigma | \Sigma_0, \nu \sim \mathcal{W}_{LM}^{-1}(\Sigma_0, \nu) \\ \lambda_0 &> 0 ; \Sigma_0 \text{ définie positive} ; \nu \geq LM \\ \mathbf{m} | \mathbf{m}_0, c_0, B &\sim \mathcal{N}_L(\mathbf{m}_0, c_0 B), \quad B | B_0, \omega \sim \mathcal{W}_L^{-1}(B_0, \omega) \\ c_0 &> 0 ; B_0 \text{ définie positive} ; \omega \geq L \end{aligned}$$

Avec ces hypothèses et sans source, la distribution marginale $p(\mathbf{x}_k | \boldsymbol{\eta})$ est alors une loi de Student multivariée :

$$\mathbf{x}_k | \boldsymbol{\eta} = \mathbf{0} \sim \text{St}_{LM} \left(\boldsymbol{\mu}_0, \frac{(\lambda_0 + 1) \Sigma_0}{\nu + 1 - LM}, \nu + 1 - LM \right)$$

Si l'on considère la distribution jointe de tous les \mathbf{x}_k , une intégration analytique de $\boldsymbol{\mu}$ et Σ mène finalement à la vraisemblance marginalisée (sur $\boldsymbol{\mu}$ et Σ) suivante :

$$p(\mathbf{x} | \boldsymbol{\eta}) = \frac{\Gamma_{LM}(\frac{\nu+K}{2}) |I_{LM} + \Sigma_0^{-1} C(\mathbf{x}, \boldsymbol{\eta})|^{-\frac{\nu+K}{2}}}{\Gamma_{LM}(\frac{\nu}{2}) (\pi^K (K \lambda_0 + 1))^{LM/2} |\Sigma_0|^{K/2}} \quad (3)$$

$$\begin{aligned} \text{avec } C(\mathbf{x}, \boldsymbol{\eta}) &= \sum_k (\mathbf{x}_k - A_k \boldsymbol{\eta})(\mathbf{x}_k - A_k \boldsymbol{\eta})^t \\ &+ \frac{1}{\lambda_0} \boldsymbol{\mu}_0 \boldsymbol{\mu}_0^t - (K + \lambda_0^{-1}) \mathbf{h}(\boldsymbol{\eta}) \mathbf{h}(\boldsymbol{\eta})^t \end{aligned}$$

$$\text{et } \mathbf{h}(\boldsymbol{\eta}) = \frac{\lambda_0}{K \lambda_0 + 1} \sum_k (\mathbf{x}_k - A_k \boldsymbol{\eta}) + \frac{1}{K \lambda_0 + 1} \boldsymbol{\mu}_0$$

L'intégration de \mathbf{m} et B mène plus simplement à l'a priori marginalisé suivant : pour $\boldsymbol{\eta}$ tel que $\eta_\ell > 0 \forall \ell$

$$p(\boldsymbol{\eta}) = \frac{(\pi(c_0 + 1))^{-\frac{L}{2}} \Gamma_L(\frac{\omega+1}{2})}{\prod_{\ell=1}^L \eta_\ell \Gamma_L(\frac{\omega}{2})} \frac{|B_0|^{\frac{\omega}{2}}}{\left| B_0 + \frac{(\ln \boldsymbol{\eta} - \mathbf{m}_0)(\ln \boldsymbol{\eta} - \mathbf{m}_0)^t}{c_0 + 1} \right|^{\frac{\omega+1}{2}}} \quad (4)$$

Il est intéressant de constater que bien que $p(\boldsymbol{\eta}) \rightarrow \infty$ quand $\boldsymbol{\eta} \rightarrow \mathbf{0}$, il est possible de démontrer que $p(\boldsymbol{\eta})$ est propre.

La distribution a posteriori $p(\boldsymbol{\eta} | \mathbf{x}) \propto p(\mathbf{x} | \boldsymbol{\eta}) p(\boldsymbol{\eta})$ est alors donnée analytiquement par les équations (3,4).

3 Détecteurs bayésiens : PLR et FBF

Le test d'hypothèse consiste à comparer une hypothèse ponctuelle imbriquée $\boldsymbol{\eta} = \mathbf{0}$ avec une hypothèse composite :

$$H_0 : \boldsymbol{\eta} = \mathbf{0} \quad H_1 : \boldsymbol{\eta} \sim p(\boldsymbol{\eta})$$

où $p(\boldsymbol{\eta})$ est l'a priori marginalisé (4).

Deux familles de détecteurs Bayésiens pour ce type de problème sont :

- la Probabilité a posteriori PLR(ζ) du rapport de vraisemblance LR($\boldsymbol{\eta}$) = $p(\mathbf{x} | \boldsymbol{\eta} = \mathbf{0}) p(\mathbf{x} | \boldsymbol{\eta})^{-1}$ [6] :

$$H_0 \text{ rejeté si } \text{PLR}(\zeta_0) > p_0, \quad (5)$$

$$\text{avec } \text{PLR}(\zeta) = \Pr\{\text{LR}(\boldsymbol{\eta}) < \zeta | \mathbf{x}\} \quad (6)$$

Textuellement, ce test -intuitif mais peu connu- consiste à détecter lorsque la probabilité a posteriori que le rapport de vraisemblance privilégie l'hypothèse H_1 est grande.

- le facteur de Bayes fractionnaire FBF(b), $b \in [0, 1[$, [7] :

$$H_0 \text{ rejeté si } \text{FBF}(b) < \zeta, \quad \text{FBF}(b) = E[\text{LR}(\boldsymbol{\eta})^{1-b} | \mathbf{x}]$$

$$\text{FBF}(b) = \frac{\int p(\mathbf{x} | \boldsymbol{\eta})^b p(\boldsymbol{\eta}) d\boldsymbol{\eta} p(\mathbf{x} | \boldsymbol{\eta} = \mathbf{0})}{\int p(\mathbf{x} | \boldsymbol{\eta}) p(\boldsymbol{\eta}) d\boldsymbol{\eta} p(\mathbf{x} | \boldsymbol{\eta} = \mathbf{0})^b} \quad (7)$$

parmi lesquels $\text{FBF}(0) = p(\mathbf{x} | H_0) p(\mathbf{x} | H_1)^{-1}$ est le facteur de Bayes classique.

Ces deux tests sont appliqués sur les vraisemblance et a priori marginalisés (3,4) dérivés en section 2.

La présence de deux paramètres dans ces familles laisse un degré de liberté intéressant pour une optimisation de ces tests, ce qui sera fait en section 5.

Les valeurs de PLR et FBF (6,7) ne peuvent pas être calculées explicitement. Pour appliquer ces tests en pratique nous proposons d'utiliser une méthode d'échantillonnage MCMC en générant une unique chaîne de Markov $\boldsymbol{\eta}^{[n]} \sim p(\boldsymbol{\eta} | \mathbf{x})$, utile aussi pour obtenir des estimateurs d'intensité.

A cause de la complexité du déterminant, la distribution a posteriori ne peut être exprimée à partir de distributions conditionnelles, ce qui est pourtant nécessaire pour l'échantillonneur de Gibbs. On utilise dans la suite le «slice sampling» qui possède des propriétés de convergence satisfaisantes et est simple d'utilisation [8].

A partir de la chaîne des rapports de vraisemblance $\text{LR}(\boldsymbol{\eta}^{[n]})$, les tests sont calculés de la façon suivante :

- FBF(b) par une procédure "d'importance sampling"
- $\text{PLR}(\zeta_0)$ en lisant la distribution empirique de $\text{LR}(\boldsymbol{\eta})$, comme l'illustre la Fig. 4.

4 Illustration sur simulation réaliste

Les tests sont appliqués à la simulation physique fine présentée en partie 1 et illustrée en Fig. 1, mais avec un contraste de luminosité différent entre l'étoile et la planète et en ne s'intéressant qu'à la zone la plus interne du champ. Deux cas sont considérés : un jeu de données dit sous H_1 avec un contraste de 10^6 ($\boldsymbol{\eta}^0 = [8, 0.5] \cdot 10^{-5}$, fixé), et un jeu de données dit sous H_0 obtenu en considérant une zone du champ limitrophe au cas sous H_1 . Les données sous H_1 , de taille $L = 2$, $K = 20$ et $M = 425$, sont illustrées en Fig. 2. Les hyperparamètres sont choisis génériques ($\nu = 2M$, $\Sigma_0 = \widehat{\sigma}^2 I_{2M}$...) ou défavorables ($\mathbf{m}_0 = \ln(1000\boldsymbol{\eta}^0)$), et identiques sous H_1 et H_0 .

Fig. 3 montre les histogrammes des chaînes de Markov résultant des deux cas. Dans le cas H_1 , les tests du Facteur de Bayes (7) et de la Probabilité a posteriori du rapport de vraisemblance (6) indiquent sans ambiguïté une détection (BF = $0.04 < \zeta_0$ et $\text{PLR}(\zeta_0) = 0.94 > 0.8$ pour $\zeta_0 = 0.1$). Dans le cas H_0 , les tests indiquent -à nouveau sans ambiguïté- qu'il n'y a pas de planète (BF = 3.7 et $\text{PLR}(\zeta_0) = 0$). Pour une information plus complète, les distributions empiriques de $\text{LR}(\boldsymbol{\eta}^{[n]})$ sont présentées en Fig. 4.

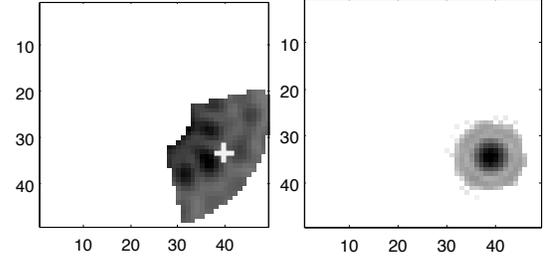


FIG. 2 – Données simulées par CAOS-SPHERE avec un contraste de 10^6 entre l'étoile et la planète. Gauche : $x_2(20)^{0.2}$. La croix indique la position de la planète. Droite : $p_2(20)^{0.2}$: le profil de la planète est supposé connu, y compris sa position.

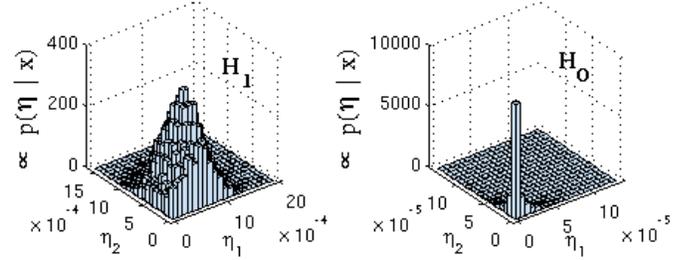


FIG. 3 – Histogrammes des deux chaînes de Markov $\boldsymbol{\eta}^{[n]} \sim p(\boldsymbol{\eta} | \mathbf{x})$ résultant des données avec planète (H_1 , gauche) et sans (H_0 , droite).

Alternativement, on utilise le test fréquentiste classique du rapport de vraisemblance généralisé sous modèle (1) en supposant en plus $\Sigma = \sigma^2 I_{LM}$. Son expression est donnée par

$$H_0 \text{ rejeté si } \text{GLRT} < \zeta, \quad \text{GLRT} = \left(\frac{\widehat{\sigma}_{H_1}^2}{\widehat{\sigma}_{H_0}^2} \right)^{\frac{KLM}{2}} \quad (8)$$

$$\text{où } \widehat{\sigma}_{H_i}^2 = \frac{\sum_k \|\mathbf{x}_k - A_k \widehat{\boldsymbol{\eta}}_{H_i} - \widehat{\boldsymbol{\mu}}_{H_i}\|^2}{KLM}$$

où $(\widehat{\boldsymbol{\mu}}_{H_1}^t, \widehat{\boldsymbol{\eta}}_{H_1}^t)$ et $\widehat{\boldsymbol{\mu}}_{H_0}$ minimisent un critère des moindres carrés obtenu directement à partir du modèle (1).

Le GLRT -contrairement à $\text{LR}(\boldsymbol{\eta}^{[n]})$ et BF- a toujours une valeur inférieure ou égale à 1. Ici, $\ln(\text{GLRT}) = -4350$ dans le cas des données sous H_1 et $\ln(\text{GLRT}) = -1300$ sous H_0 . Comme il n'est pas numériquement possible de générer un grand nombre de jeux de données simulées finement, il n'est pas possible de relier numériquement le seuil du GLRT à sa Probabilité de Fausse Alarme (PFA). Le modèle (1) n'est de plus pas identiquement distribué, ce qui empêche d'utiliser les résultats classiques sur la distribution asymptotique du GLRT. Il est donc difficile de choisir le seuil ζ , ce qui est en soi un problème quant à l'applicabilité pratique du GLRT. Le GLRT appliqué sur des zones proches mais distinctes des cas précédents -résultats non présentés ici par manque d'espace- indique que le GLRT discrimine de toute façon difficilement H_0 et H_1 .

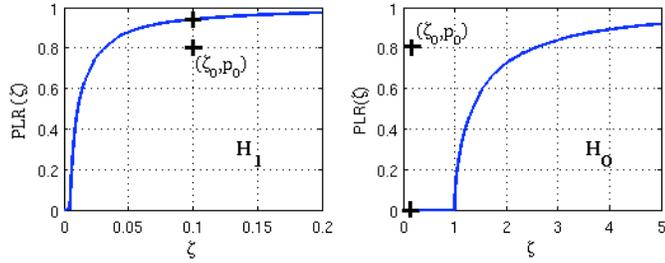


FIG. 4 – Distributions cumulatives empiriques a posteriori de LR, tracées à partir des chaînes $\eta^{[n]}$ présentées en Fig. 3. On lit à gauche que H_0 est rejeté par le test PLR pour $\zeta_0 = 0.1$ et $p_0 = 0.8$ car alors $\text{PLR}(\zeta_0) = \Pr\{\text{LR}(\eta) < \zeta_0 | \mathbf{x}\} > p_0$.

5 Courbes CORE et optimisation sur simulations simples

On utilise la courbe CORE -outil fréquentiste- pour comparer et optimiser numériquement les performances des familles de tests Bayésiens. PFA et PD sont estimées par des intégrations de Monte Carlo classiques à partir d'un grand nombre de jeux de données \mathbf{x} . L'application des tests Bayésiens requérant pour chaque \mathbf{x} la génération d'une chaîne de Markov $\eta^{[n]}$ de grande taille, les données doivent être de taille restreinte.

En pratique les données sont donc simulées à partir de la vraisemblance marginalisée (3) pour $KLM = 80$. Par simplicité, les données sous H_1 sont caractérisées par une source d'intensité $\eta = \eta^0$ fixée.

Fig. 5 illustre les courbes CORE obtenues pour des valeurs de paramètre fixées : p_0 ou ζ_0 pour le PLR (Eq. 6) et b pour le FBF (Eq. 7). Elle montre qu'ici le facteur de Bayes classique (FBF(0)) est dépassé uniformément par le PLR et le FBF paramétrés de façon standard en fixant respectivement $p_0 = 0.8$, $\zeta_0 = 1$ et $b = 0.3$. Pour PFA = 0.1, les performances du PLR dépassent celles du Bayes Factor de 15%. Fig. 6 montre la valeur de seuil ζ_0 à choisir pour atteindre une PFA fixée pour le PLR et le FBF lorsque l'autre paramètre est fixé.

Une optimisation numérique des familles de tests précédentes laissant les deux paramètres libres ($\lambda = (\zeta_0, p_0)$ pour PLR et (b, ζ_0) pour FBF) est possible. Par exemple, dans le cas du PLR, pour chaque PFA fixée on détermine le paramètre $\lambda(\text{PFA})$ qui maximise PD. Les deux courbes CORE associées à ces tests optimaux sont représentées sur la Fig. 5. On constate que les tests à paramètres fixés montrés précédemment ont des performances très proches des performances optimales pour l'intervalle de PFA et l'intensité η_0 considérés.

Fig. 5 indique finalement à nouveau les mauvaises performances du test du GLRT (8) où Σ était supposée -à tort- proportionnelle à l'identité.

Références

[1] J. Schneider, "Extrasolar planets encyclopaedia." [Online]. Available : <http://exoplanet.eu/>

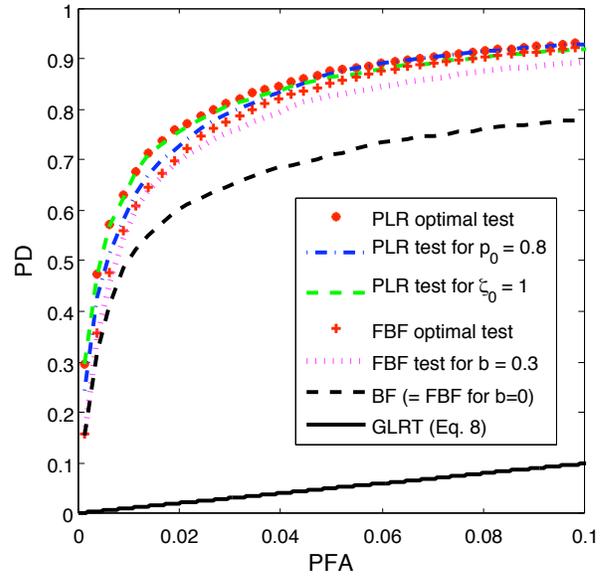


FIG. 5 – Courbes CORE pour le PLR et le FBF pour différentes paramétrisations, et courbe CORE du GLRT sous hypothèse $\Sigma = \sigma^2 I_{LM}$.

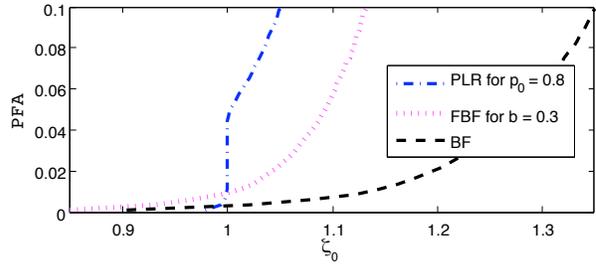


FIG. 6 – Relation entre la PFA et le seuil ζ_0 pour des tests du PLR et du FBF.

- [2] M. Carbillat, A. Boccaletti, C. Thalmann *et al.*, "The software package SPHERE : a numerical tool for end-to-end simulations of the VLT instrument SPHERE," in *Adaptive Optics Systems*, N. Hubin, C. E. Max, and P. L. Wizinowich, Eds. SPIE, 2008.
- [3] L. Mugnier, J.-F. Sauvage, T. Fusco, and G. Rousset, "Multi-channel planet detection algorithm for angular differential imaging," in *Optical Society of America*, ser. Adaptive Optics : Analysis and Methods, B. Ellerbroek, Ed., 2007.
- [4] I. Smith, A. Ferrari, and M. Carbillat, "Detection of a moving source in speckle noise. application to exoplanet detection," *IEEE Transactions on Signal Processing*, 2008.
- [5] C. P. Robert, *The Bayesian Choice*. Springer, February 2001.
- [6] M. Aitkin, "The calibration of P-values, posterior Bayes factors and the AIC from the posterior distribution of the likelihood," *Statistics and Computing*, vol. 7, no. 4, pp. 253–261, 1997.
- [7] A. O'Hagan, "Fractional bayes factors for model comparison," *Journal of the Royal Statistical Society*, 1995.
- [8] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer-Verlag, 1999.