

1

Validation des mocks

Ce chapitre a pour vocation de présenter l’analyse menée sur les mocks, afin de valider leur construction et le choix des différents paramètres. Nous présentons d’abord le calcul des diverses fonctions de corrélations, puis l’analyse des 100 réalisations produites. Pour faire cette analyse, nous utilisons le code `picca` qui permet de calculer les champs δ , les fonctions de corrélations, les matrices de distorsion, les matrices de covariance ainsi que de réaliser l’ajustement du modèle.

1 Les estimateurs

Nous présentons ici les estimateurs utilisés pour calculer la fonction d’auto-corrélation du $\text{Ly}\alpha$, la fonction de corrélation croisée $\text{Ly}\alpha$ -QSO, ainsi que la fonction d’auto-corrélation d’objets ponctuels tels les quasars. Dans le cas des objets ponctuels, la seule information nécessaire au calcul de la fonction de corrélation est le catalogue fournissant pour chaque objet l’ascension droite, la déclinaison et le redshift. Pour le calcul des fonctions de corrélation impliquant le $\text{Ly}\alpha$, il est nécessaire de calculer au préalable le champ δ_F (équation ??). Nous distinguons deux cas : le cas où nous analysons les *raw mocks* et le cas où nous analysons les *cooked mock* ou les données. Les raw mocks (les mocks bruts, non préparés) désignent les mocks avant d’avoir tourné le code `quickquasars`. Les fonctions de corrélation sont alors calculées en utilisant directement la fraction de flux transmise F donnée dans les fichiers de transmissions. Etant donné que pour chaque forêt, nous avons directement accès à F , le champ δ_F est donné par

$$\delta_F = \frac{F}{\overline{F}} - 1. \quad (1.1)$$

Les cooked mocks quant à eux désignent les mocks après avoir appliqué `quickquasars` à chaque forêt. Dans ce cas, comme pour les données, la fraction de flux transmise F n’est pas accessible. Il faut alors utiliser la procédure d’ajustement du continuum décrite dans la section ?. Une fois le champ δ_F calculé, nous pouvons calculer les différentes fonctions de corrélation.

1.1 L’auto-corrélation $\text{Ly}\alpha \times \text{Ly}\alpha$

Comme expliqué dans le chapitre d’introduction, la fonction d’auto-corrélation du champ d’absorption F du $\text{Ly}\alpha$ est définie comme

$$\xi_{\text{Ly}\alpha}(\vec{r}) = \langle \delta_F(\vec{r}') \delta_F(\vec{r} + \vec{r}') \rangle. \quad (1.2)$$

Elle peut-être mise sous la forme

$$\xi(\vec{r}) = \langle \delta_i \delta_j \rangle, \quad (1.3)$$

où $\langle . \rangle$ désigne la moyenne sur tous les pixels i et j qui vérifient $\vec{r}_{ij} = \vec{r}$. Afin de calculer $\xi_{\text{Ly}\alpha}$, nous créons une grille en $(r_{\parallel}, r_{\perp})$. Les bins mesurent $4 h^{-1}$ Mpc dans chaque direction. Cette taille de bin permet à la fois de regrouper suffisamment de pixels pour avoir une bonne statistique dans chaque bin, mais aussi d’avoir suffisamment de bins pour résoudre correctement la région du pic BAO (#prov est-ce que c’est vrai ? toujours la même quantité d’info, donc ça améliore pas le fit ?). Une fois cette grille construite, nous transformons les coordonnées (α, δ, z) des pixels δ_F en distances (#prov donner les transformations géométriques ?). Puis, pour chaque bin A de la grille en $(r_{\parallel}, r_{\perp})$, nous considérons toutes les paires de pixels dont la distance de séparation se trouve dans ce bin A . La fonction de corrélation dans ce bin est alors donnée par

$$\xi_A = \frac{\sum_{(i,j) \in A} w_i w_j \delta_i \delta_j}{\sum_{(i,j) \in A} w_i w_j}, \quad (1.4)$$

où w_i est le poids associé au pixel i . Ces poids sont définis dans l’équation ?. La correction de la dépendance en redshift du biais du $\text{Ly}\alpha$ permet de donner plus de poids aux pixels à grand redshift,

où l'amplitude de la fonction de corrélation est plus importante. Le calcul de la fonction de corrélation est donc effectué à l'aide d'une double boucle sur les pixels. Afin de réduire le temps de calcul, nous calculons la fonction de corrélation uniquement pour les paires de pixels pour lesquelles r_{\parallel} et r_{\perp} sont inclus dans $[0; 200]h^{-1}$ Mpc. La fonction de corrélation est donc calculée dans $50 \times 50 = 2500$ bins. Les paires formées par des pixels provenant de la même forêt sont exclues du calcul afin d'éviter que les erreurs sur l'ajustement du continuum biaisent la mesure de la fonction de corrélation.

L'analyse Ly α des données complète d'eBOSS (?) utilise deux fonctions de corrélation distinctes : les fonctions de corrélation Ly α (Ly α) \times Ly α (Ly α) et Ly α (Ly α) \times Ly α (Ly β). La région Ly β ne fournit pas suffisamment de pixels pour pouvoir calculer la fonction de corrélation Ly α (Ly β) \times Ly α (Ly β) et y détecter le pic BAO. L'utilisation de ces deux fonctions de corrélation permet de mesurer la position du pic BAO avec une plus grande précision. Dans ce manuscrit, nous nous intéressons à la mesure de $b_{\text{Ly}\alpha}$ et $\beta_{\text{Ly}\alpha}$ afin de construire correctement nos mocks. Pour limiter les potentielles systématiques, nous considérons donc uniquement la fonction de corrélation Ly α (Ly α) \times Ly α (Ly α). Le graphique de droite de la figure ?? présente les distributions pondérées en redshift des paires Ly α (Ly α) \times Ly α (Ly α) et Ly α (Ly α) \times Ly α (Ly β). L'analyse des données DR16 que nous présentons ici considère donc uniquement la distribution indiquée en orange.

1.2 La corrélation croisée Ly α \times QSO

Nous donnons dans cette section l'estimateur de la fonction de corrélation croisée Ly α \times QSO. De la même manière que précédemment, le calcul s'effectue dans des bins en $(r_{\parallel}, r_{\perp})$. La fonction de corrélation dans le bin A est donnée par

$$\xi_A = \frac{\sum_{(i,j) \in A} w_i w_j \delta_i}{\sum_{(i,j) \in A} w_i w_j}, \quad (1.5)$$

où l'indice i court sur les pixels des forêts et j sur les quasars pour lesquels la distance de séparation est comprise dans le bin A . Comme précédemment, les paires ij où le pixel i appartient à la forêt du quasar j sont rejetées. Les poids w_j associés aux quasars, similairement au Ly α , favorisent les quasars à plus grand redshift. Ces poids sont définis comme

$$w_j = \left(\frac{1 + z_j}{1 + 2.25} \right)^{\gamma_{QSO}}, \quad (1.6)$$

où $\gamma_{QSO} = 1.44 \pm 0.08$ (?). Dans le calcul de la fonction de corrélation croisée Ly α \times QSO, nous nous restreignons aux paires pour lesquelles $r_{\perp} \in [0; 200]h^{-1}$ Mpc. Contrairement à l'auto-corrélation du Ly α , la fonction de corrélation croisée n'est pas symétrique en r_{\parallel} . Nous la calculons donc dans les bins pour lesquels $r_{\parallel} \in [-200; 200]h^{-1}$ Mpc. La fonction de corrélation croisée est donc calculée dans $50 \times 50 = 2500$ bins.

1.3 Le spectre de puissance à une dimension

Comme expliqué dans le chapitre précédent, nous ajustons le spectre de puissance P_{miss} appliqué à δ_s de façon à obtenir un spectre de puissance à une dimension P^{1D} en accord avec les données. Nous présentons donc maintenant la mesure de ce spectre de puissance. Pour chaque forêt, nous appliquons une transformation de Fourier au champ δ_F afin d'obtenir le champ δ_k . Puis, le spectre de puissance de cette forêt est obtenu comme

$$P^{1D}(k) = |\delta_k^2|. \quad (1.7)$$

Nous répétons cette procédure pour toutes les forêts, puis le spectre de puissance total est obtenu comme la moyenne du spectre de puissance de chaque forêt.

En ce qui concerne les données, la mesure du spectre de puissance à une dimension est plus complexe. Un certain nombre d'effets liés à la mesure doivent être pris en compte. Le bruit, par exemple, doit être estimé afin d'être pris en compte pour ne pas fausser l'estimation du spectre de puissance. Cette analyse est détaillée dans ?.

1.4 La fonction de corrélation à une dimension

Important / nécessaire ?

2 Les matrices de distorsion

L'ajustement du continuum nécessaire au calcul du champ δ_F dans les données et les cooked mocks biaise le champ mesuré. Cependant, grâce à la transformation (équation ??) décrite dans la section ??), l'effet sur la fonction de corrélation peut-être pris en compte. L'idée est la suivante : modéliser la distorsion induite par l'ajustement du continuum et par la transformation ?? sur la fonction de corrélation, appliquer cette distorsion au modèle, puis ajuster le modèle "distordu" aux données. L'ajustement du continuum et la transformation ?? induisent des corrélation le long de la ligne de visée. Au premier ordre, nous pouvons considérer que chaque δ_F d'une forêt est une combinaison linéaire de tous les δ_F de cette forêt. La fonction de corrélation distordue dans le bin A peut alors être reliée à la vraie fonction de corrélation comme

$$\xi_{distortion}(A) = \sum_B D_{AB} \xi_{vraie}(B), \quad (1.8)$$

où D_{AB} est appelée la *matrice de distorsion*. Celle-ci s'exprime en fonction du projecteur η_{ij}^q défini dans l'équation ??. Pour l'auto-corrélation, elle est défini comme

$$D_{AB} = \frac{1}{W_A} \sum_{(i,j) \in A} w_i w_j \left(\sum_{(i',j') \in B} \eta_{ii'} \eta_{jj'} \right) \quad (1.9)$$

où $W_A = \sum_{(i,j) \in A} w_i w_j$ est le poids du bin A . Pour la corrélation croisée, la matrice de distorsion est donnée par

$$D_{AB} = \frac{1}{W_A} \sum_{(i,j) \in A} w_i w_j \left(\sum_{(i',j') \in B} \eta_{ii'} \right). \quad (1.10)$$

Comme précédemment, les indices i correspondent aux pixels des forêts, et j aux quasars. A cause de la double somme, le calcul de la matrice de distorsion est très long. Afin de rendre possible l'estimation de cette dernière, le calcul est fait sur 1 % des paires, tirées au hasard (#prov Bautista 2017 montre que c'est OK avec 5%, est-ce que y a une etude qui montre que c'est ok avec 1% ? Dans DR14, ils utilisent 5% je pense).

La figure ?? (#prov faire la figure) montre l'effet de la distorsion sur l'auto-corrélation $\text{Ly}\alpha \times \text{Ly}\alpha$ dans les mocks. La fonction de corrélation est présentée dans quatres gammes en μ . La courbe noire indique la fonction de corrélation calculée sur les raw mocks, et la courbe bleu la fonction de #prov Refaire la figure 11 de Bautista : CF des raw mocks (stack) + CF sur les cooked mocks (stack) avec le fit sur les raw + fit * DM

La matrice de distorsion est un objet uniquement géométrique. Son calcul est indépendant du champ δ_F . Elle ne dépend uniquement de la géométrie du relevé (#prov comment ?) et de la distribution des poids. La figure ?? montre la différence entre #prov montre la difference du fit d'une réa avec sa DM et avec une autre DM et/ou le stack de 10 avec les 10 DM ou dix fois la meme. Est-ce que les DM eboss-0.0 et eboss-0.2 sont différentes ? Ainsi, il n'est pas nécessaire de calculer la matrice

de distorsion pour les 100 réalisations des mocks. Dans l'analyse présentée dans la suite de ce chapitre, nous calculons la matrice de distorsion pour l'auto-corrélation et pour la corrélation croisée une seule fois (#prov ou une fois pour chaque run de quickquasars?). L'ajustement de chaque fonction de corrélation utilise une de ces deux matrices de distorsion.

3 Les matrices de covariance

Afin de réaliser l'ajustement de chaque fonction de corrélation, nous avons besoin de calculer les matrices de covariance associées à ces fonctions de corrélation. La covariance de la fonction de corrélation ξ dans le bin A et de ξ dans le bin B est définie comme

$$C_{AB} = \langle \xi_A \xi_B \rangle - \langle \xi_A \rangle \langle \xi_B \rangle . \quad (1.11)$$

De cette matrice de covariance, nous définissons la matrice de corrélation comme

$$Corr_{AB} = \frac{C_{AB}}{\sqrt{C_{AA}C_{BB}}} . \quad (1.12)$$

La matrice de corrélation donne la corrélation, comprise dans $[-1;1]$, d'un bin A avec un bin B . Afin d'estimer la matrice de covariance, le relevé est divisé en HEALPix pixels, en utilisant `nside` = 16. Cette résolution produit des pixels d'une taille sur le ciel de $3,7 \times 3,7 = 13,4 \text{ deg}^2$, correspondant à $250 \times 250 (h^{-1} \text{ Mpc})^2$ à un redshift $z = 2,33$. Ces sous-échantillons sont suffisamment grands pour pouvoir négliger les corrélations entre différents HEALPix pixels et ainsi estimer la matrice de covariance comme la variance d'un sous-échantillon à un autre. La matrice de covariance est donc calculée comme

$$C_{AB} = \frac{1}{W_A W_B} \sum_s W_A^s W_B^s (\xi_A^s \xi_B^s - \xi_A \xi_B) , \quad (1.13)$$

où s est un sous-échantillon, et W_A^s les poids du bin A de ce sous-échantillon. Les principaux éléments de cette matrice sont les éléments diagonaux : la variance dans chaque bin. Les éléments non-diagonaux, les covariances entre deux bins distincts, sont faibles (#prov donner comme `Var_A` est modélisé? Donner comment on modélise les éléments non diag?). Leur estimation est bruitée. La matrice de covariance est donc lissée après avoir été estimée. En ce qui concerne l'auto-corrélation, la matrice de covariance possède 2500×2500 bins. Pour la corrélation croisée, elle en possède 5000×5000 .

4 Modélisation des fonctions de corrélation

Dans cette section, nous présentons les modèles utilisés pour ajuster les fonctions de corrélation $\text{Ly}\alpha \times \text{Ly}\alpha$ et $\text{Ly}\alpha \times \text{QSO}$. Nous présentons d'abord les modèles utilisés dans l'analyse des données DR16, puis nous donnons les modèles utilisés pour analyser les mocks.

4.1 Modélisation des données

Pour l'analyse des données DR16, nous utilisons le modèle décrit dans ?. L'analyse décrite dans cette étude est une analyse BAO : l'auto-corrélation et la corrélation croisée sont modélisées de façon à mesurer au mieux les paramètres BAO α_{\parallel} et α_{\perp} . Pour ce faire, le modèle est séparé en deux composantes. La première, ξ_{smooth} , correspond à la forme globale de la fonction de corrélation. La seconde, ξ_{peak} , correspond au pic BAO. C'est cette seconde composante qui dépend des paramètres BAO :

$$\xi(r_{\parallel}, r_{\perp}, \alpha_{\parallel}, \alpha_{\perp}) = \xi_{\text{smooth}}(r_{\parallel}, r_{\perp}) + \xi_{\text{peak}}(\alpha_{\parallel} r_{\parallel}, \alpha_{\perp} r_{\perp}) . \quad (1.14)$$

Cette séparation est opérée au niveau du spectre de puissance. Le modèle de la fonction de corrélation est ensuite obtenue à l'aide d'une transformation de Fourier.

Le spectre de puissance

Le spectre de puissance utilisé dans la modélisation de la fonction de corrélation des traceurs i et j s'exprime comme

$$P(\vec{k}) = b_i b_j (1 + \beta_i \mu_k^2) (1 + \beta_j \mu_k^2) P_{\text{QL}}(\vec{k}) F_{\text{NL}}(\vec{k}) G(\vec{k}) . \quad (1.15)$$

Les termes $b_i(1 + \beta_i \mu_k^2)$ et $b_j(1 + \beta_j \mu_k^2)$ sont les facteurs de Kaiser (équation ??) relatifs aux traceurs i et j . P_{QL} est le spectre de puissance *quasi-linéaire*. Il est découpé en deux composantes P_{smooth} et P_{peak} comme

$$P_{\text{QL}}(\vec{k}, z) = P_{\text{smooth}}(k, z) + \exp\left(-\frac{k_{\parallel}^2 \Sigma_{\parallel}^2 + k_{\perp}^2 \Sigma_{\perp}^2}{2}\right) P_{\text{peak}}(k, z) . \quad (1.16)$$

P_{smooth} est le spectre de puissance linéaire, sans les BAO. Il est construit à partir du spectre de puissance linéaire P_{L} donné par Camb, puis les BAO sont retirées en utilisant la technique *side-band* décrite dans ?. Le spectre de puissance P_{peak} est alors obtenu comme la différence $P_{\text{L}} - P_{\text{smooth}}$: il contient uniquement les oscillations dues aux BAO présentes dans le spectre de puissance linéaire. Le terme exponentiel devant P_{peak} , paramétré par Σ_{\parallel} et Σ_{\perp} , prend en compte l'élargissement non linéaire du pic BAO. Nous utilisons $\Sigma_{\parallel} = 6,42 h^{-1} \text{ Mpc}$ et $\Sigma_{\perp} = 3,26 h^{-1} \text{ Mpc}$. Le terme F_{NL} prend en compte les non linéarités aux petites échelles. Nous distinguons $F_{\text{NL}}^{\text{auto}}$ et $F_{\text{NL}}^{\text{cross}}$. Pour l'auto-corrélation, les effets non linéaires proviennent de l'élargissement thermique, des vitesses particulières et de la croissance des structures non linéaire. Comme lors de la modélisation du P^{1D} , nous utilisons le modèle décrit dans ?. Nous avons donc $F_{\text{NL}}^{\text{auto}}(k, \mu) = D(k, \mu)$, où D est défini dans l'équation ??. Les paramètres utilisés sont une interpolation à $z = 2,334$ de ceux donnés dans la section "Planck" de la table 7 de ?. Pour la corrélation croisée, l'effet dominant est dû aux vitesses non linéaires des quasars. Cet effet est modélisé par une lorentzienne :

$$F_{\text{NL}}^{\text{cross}}(k_{\parallel}) = \frac{1}{1 + (k_{\parallel} \sigma_v)^2} , \quad (1.17)$$

où l'inverse de la demi-largeur à mi-hauteur σ_v est un paramètre libre. L'effet dû aux erreurs statistiques sur la mesure du redshift des quasars étant confondu avec l'effet des vitesses non linéaires des quasars, il est aussi pris en compte par le terme $F_{\text{NL}}^{\text{cross}}$. Enfin, le terme $G(\vec{k})$ prend en compte l'effet du binning utilisé lors du calcul de la fonction de corrélation. Il est défini comme

$$G(\vec{k}) = \text{sinc}\left(\frac{k_{\parallel} R_{\parallel}}{2}\right) \text{sinc}\left(\frac{k_{\perp} R_{\perp}}{2}\right) , \quad (1.18)$$

avec R_{\parallel} et R_{\perp} la largeur des bins, soit $4 h^{-1} \text{ Mpc}$.

Modélisation des HCD