



Facultad de Matemática,
Astronomía, Física y
Computación



Universidad
Nacional
de Córdoba

GALAXIAS Y ESTRUCTURAS EN VACÍOS CÓSMICOS

POR

LIC. FEDERICO DÁVILA KURBÁN

Presentada ante la Facultad de Matemática, Astronomía, y Física
como parte de los requerimientos para obtener el título de Doctor en
Astronomía de la

UNIVERSIDAD NACIONAL DE CÓRDOBA, ARGENTINA

Julio 2022

Director: Dr. Marcelo Lares

Codirector: Dr. Diego García Lambas

Tribunal Especial:

Titulares:

Dr. Nelson Padilla (OAC-UNC)

Dr. Facundo Gomez (ULS, Chile)

Dra. Carolina Chavero (OAC-UNC)

Suplentes:

Dra. Andrea Ahumada (OAC-UNC)

Dra. Patricia Tissera (UAB, Chile)



Esta obra está bajo una licencia [Creative Commons](https://creativecommons.org/licenses/by-sa/4.0/) “Atribución-CompartirIgual 4.0 Internacional”.

Galaxias y estructuras en vacíos cósmicos

Lic. Federico Dávila Kurbán

Tesis Doctoral

Facultad de Matemática, Astronomía y Física

Universidad Nacional de Córdoba

Resumen: El análisis de la distribución espacial de galaxias en gran escala ha sido fundamental para entender la evolución del Universo. En las escalas más grandes, esta distribución demuestra características geométricas, usualmente referidas conjuntamente como la red cósmica, compuesta de paredes, nodos, filamentos, y vacíos. Los vacíos cósmicos son las regiones menos densas y, por lo tanto, uno de los ambientes más extremos en los cuales las galaxias pueden formarse y evolucionar con una tasa de interacciones menor que en otros entornos. Por lo tanto, las galaxias y su distribución en estos entornos *subdensos* constituyen la mejor aproximación a las propiedades pristinas de las mismas. Sin embargo, el estudio estadístico de propiedades de las galaxias y las estructuras que forman en gran escala en estos entornos subdensos se ve naturalmente limitado por el pequeño tamaño de las muestras. En esta Tesis se proponen nuevos enfoques estadísticos que aplicamos en simulaciones numéricas y que apuntan a mejorar las herramientas de análisis, a saber: métodos del estudio de la agrupación de galaxias que reducen el error de los estadísticos conocidos o reducen su tiempo de cómputo; estadísticos para el análisis de las orientaciones de las galaxias que apuntan a usar toda la información disponible y con una robusta determinación de su significancia estadística; y un estudio de la agrupación de galaxias empleando un método que, a diferencia de los estadísticos usuales, analiza regiones de exclusión de galaxias y por lo tanto resulta natural su empleo en regiones subdensas. Estos enfoques y herramientas, además, implican un gran provecho de la información

disponible con el advenimiento de simulaciones de mayor resolución y catálogos más numerosos.

Primero, abordamos la práctica más general del estudio del agrupamiento de galaxias en el Universo e intentamos mejorarlo. Las herramientas más usadas para caracterizar la estructura del Universo a gran escala son estadísticos bipuntuales, tales como el espectro de potencia, $P(\mathbf{k})$, y su transformada de Fourier, la función de correlación bipuntual, $\xi(\mathbf{s})$. En el espacio de configuraciones, los estimadores de $\xi(\mathbf{s})$ cuantifican el exceso de probabilidad de encontrar un par de galaxias a un dado vector de separación \mathbf{s} con respecto a una distribución de referencia sin agrupamiento, usualmente denominada “catálogo aleatorio”. Estimaciones de alta precisión hacen uso de grandes catálogos aleatorios lo cual implica un alto costo computacional. Proponemos reemplazar los catálogos aleatorios estándares por distribuciones puntuales tipo *glass* o *catálogos glass* cuyo espectro de potencia $P(k) \propto k^4$ exhibe significativamente menos potencia en escalas mayores que la separación media entre partículas, que una distribución Poisson (cuyo $P(k)$ es constante) con el mismo número de puntos. Mostramos que estas distribuciones pueden ser obtenidas iterativamente aplicando la técnica de reconstrucción de Zeldovich comúnmente usada en estudios de oscilaciones bariónicas acústicas. Proveemos una versión modificada del ampliamente usado estimador Landy–Szalay de la función de correlación adaptada al uso de catálogos *glass* y comparamos su desempeño con resultados obtenidos usando catálogos aleatorios. Nuestros resultados muestran que las distribuciones tipo *glass* no introducen ningún sesgo respecto a los resultados obtenidos usando distribuciones Poisson. En escalas mayores a la separación media entre partículas de los catálogos *glass*, el estimador modificado reduce significativamente la varianza de la función de correlación bipuntual con respecto a los resultados estándares obtenidos con el estimador Landy–Szalay usando la misma cantidad de puntos. El tamaño del catálogo *glass* que se requiere para llegar a una dada precisión en la función de correlación es significativamente menor que el necesario con un catálogo aleatorio. Su uso podría reducir drásticamente el costo computacional, manteniendo una alta precisión, del análisis del agrupamiento de galaxias en relevamientos futuros.

La naturaleza del agrupamiento de galaxias depende de numerosos efectos a pequeñas y grandes escalas, tales como los parámetros cosmológicos, efectos e historia del entorno de galaxias y cúmulos, la distribución subyacente de materia oscura, y la forma en la que las componentes luminosas y oscuras del Universo

se acoplan y evolucionan. Al estudiar los órdenes bajos y altos del agrupamiento de galaxias, se espera arrojar luz sobre los procesos físicos de los cuales depende el agrupamiento. A pesar de su utilidad, la función de correlación bipuntual sólo provee una descripción completa del agrupamiento en el caso de una distribución gaussiana. Un recuento más exhaustivo del agrupamiento debe incluir funciones de correlación de altos órdenes, sin embargo estos suelen ser difíciles de extraer. A la luz de esto, investigadores han explorado otros estadísticos de agrupamiento con el fin de obtener información de alto orden acerca de la distribución de galaxias, en particular estadísticos de vacío tales como la *función de probabilidad de vacío* (VPF, por sus siglas en inglés), que mide la probabilidad de encontrar cero galaxias en un dado radio. Este enfoque ha resultado provechoso en el sentido de que los resultados están bien fundamentados por un marco teórico sólido que relaciona la distribución de vacíos (i.e., la VPF) a la distribución de alto orden de galaxias.

En la labor de caracterizar la estructura interna de los vacíos cósmicos, hemos estudiado la VPF dentro y fuera de vacíos identificados en la simulación TNG300-1, en el espacio real y en el espacio de redshift. Elegimos esta simulación ya que es la más grande y con mayor resolución de las simulaciones TNG. Se ha mostrado previamente que las galaxias en el espacio de redshift siguen el modelo estadístico denominado “binomial negativo”, pero este acuerdo no se traslada al espacio real. Esto significa, bajo la formulación de la VPF, que en el espacio de redshift hay una aparente relación de escala entre la mencionada función de correlación bipuntual (el orden más bajo de agrupación) y los órdenes de agrupación superiores. Nuestros resultados indican, sin embargo, que las galaxias dentro de los vacíos cósmicos en el espacio real, siguen el modelo binomial igual que las galaxias en el espacio de redshift; i.e. estudiando galaxias en vacíos cósmicos recuperamos el escalamiento jerárquico detectado en el espacio de redshift.

Como hemos establecido, las galaxias son de interés primordial en el estudio del Universo. Analizando las distribuciones espaciales y espectrales de su luz en combinación con sus propiedades dinámicas, se pueden inferir los procesos físicos que gobiernan la formación de galaxias e intentar explicar la abundancia de tipos de galaxias y estructuras observadas. En particular, son de interés las orientaciones y elipticidades de las imágenes galácticas. Orientaciones preferenciales, o alineamientos, de galaxias entre sí, con las estructuras de materia subyacentes y con la mencionada red cósmica son elementos cruciales para avanzar hacia un

entendimiento más completo de la gravedad, la naturaleza de la materia, y la formación de estructuras en el Universo. Proponemos estadísticos bien comportados para estudiar la señal de alineamiento de los momentos angulares, o *spines*, con respecto al centro de los vacíos, de galaxias ubicadas en zonas subdensas alrededor de vacíos identificados en la simulación TNG300-1. Exploramos esta señal en distintas poblaciones de galaxias, variando su distancia al centro del vacío, su masa, la norma de su spin, la densidad local, y su velocidad. Encontramos una fuerte tendencia ($> 9\sigma$) de alineamiento perpendicular a la dirección al centro del vacío en galaxias masivas, con alto spin, y baja velocidad radial, en distancias entre 0.9 y 1.4 radios de vacío. Además, encontramos que en estos ambientes subdensos, la densidad local es irrelevante a la señal de alineamiento, mientras que el mayor impacto en la señal lo tiene la velocidad radial: las galaxias que pierden momento lineal con la expansión del vacío parecen estar fuertemente alineadas perpendicularmente a la dirección al centro del mismo. Un mayor entendimiento del alineamiento intrínseco será crucial para interpretar estudios de *weak lensing*, y para capitalizar los grandes relevamientos de *shear* cósmico como Euclid o LSST.

En total, proponemos métodos y estadísticos nuevos para estudiar el agrupamiento de galaxias por un lado, y las orientaciones con respecto a los vacíos por otro, y finalmente utilizamos medidas no estándares de agrupamiento para caracterizar la distribución de galaxias adentro de los vacíos, comparado con la distribución en el resto de la red cósmica. Hemos aprendido que los vacíos son un entorno único en los cuales la estructura puede formarse y evolucionar, y que pueden jugar un rol crucial en otorgar a las galaxias de su entorno alineamiento y momento angular.

Abstract: The analysis of the large-scale spatial distribution of galaxies has been instrumental in understanding the evolution of the Universe. On the largest scales it displays striking geometrical features, usually referred to as the cosmic web, composed of walls, clusters, filaments, and voids. Cosmic voids are the most underdense regions and thus one of the most extreme environments where galaxies can form and evolve with a less frequent interaction rate than in the other environments. Therefore, galaxies and their distribution in these *subdense* environments constitute the best approximation to their pristine properties. However, the statistical study of the properties of galaxies and the large-scale structures they form in these subdense environments are naturally limited by the small sample sizes. This Thesis proposes new statistical approaches that are applied in numerical simulations and aim to improve the tools of analysis, namely: methods of studying the clustering of galaxies that reduce the error of the statistics or their computational time; statistics for the analysis of galaxy orientations that aim at using all the information available and with a robust determination of their statistical significance; and finally a study of the clustering of galaxies using a method that, unlike the more usual statistics, analyses the regions of exclusion of galaxies and is therefore a natural choice for subdense environments. These approaches and tools also imply a great benefit from the amount of information available with upcoming simulations with greater resolution and larger surveys.

First, we take on the more general practice of the study of galaxy clustering throughout the Universe and attempt to improve upon it. The most commonly used tools to characterize the large-scale structure of the Universe are two-point statistics such as the power spectrum, $P(\mathbf{k})$, and its Fourier transform, the two-point correlation function $\xi(\mathbf{s})$. In configuration space, the estimators of $\xi(\mathbf{s})$ quantify the excess probability of finding a pair of galaxies at a given separation vector \mathbf{s} with respect to a reference distribution with no clustering, often referred to as the “random catalogue”. High-accuracy estimates require the use of large random catalogues, which imply a high computational cost. We propose to replace the standard random catalogues by glass-like point distributions or *glass catalogues* whose power spectrum $P(k) \propto k^4$ exhibits significantly less power on scales larger than the mean inter-particle separation than a Poisson distribution (whose $P(k)$ is constant) with the same number of points. We show that these distributions can be obtained by iteratively applying the technique of Zeldovich reconstruction commonly used in studies of the baryon acoustic oscillations. We provide a modi-

fied version of the widely used Landy-Szalay estimator of the correlation function adapted to the use of glass catalogues and compare its performance with the results obtained using random samples. Our results show that glass-like samples do not add any significant bias with respect to the results obtained using Poisson distributions. On scales larger than the mean inter-particle separation of the glass catalogues, the modified estimator leads to a significant reduction of the variance of the two-point correlation function with respect to the standard Landy-Szalay results with the same number of points. The size of the glass catalogue required to achieve a given accuracy in the correlation function is significantly smaller than when using random samples. Their use could help to drastically reduce the computational cost of configuration-space clustering analysis of future surveys while maintaining high-accuracy.

The nature of galaxy clustering is dependent on many large- and small-scale effects, such as the cosmological parameters, galaxy and cluster environmental effects and history, the underlying dark matter distribution, and the way in which the dark and luminous components of the Universe couple and evolve. By probing the lower and higher orders of galaxy clustering, one thus hopes to shed light on those physical processes on which the clustering is dependent. And in spite of its usefulness, the two-point correlation function only provides a complete description of clustering in the case of a Gaussian distribution. A more exhaustive account of clustering must include higher-order correlation functions, although these prove difficult to extract. In light of this, researchers have explored other clustering statistics in order to obtain high-order information about galaxy distributions, in particular void statistics such as the Void Probability Function (VPF), that measures the probability of finding zero galaxies in a given radius. This approach has proven to be useful in the sense that results are well supported by a solid theoretical framework that directly relates the distribution of voids (i.e., the VPF) to the high-order distribution of galaxies.

In the endeavor of characterizing structure inside cosmic voids, we have studied the VPF inside and outside voids identified in the TNG300-1 simulation, in real space as well as in redshift space. We chose this simulation because it is the largest simulation and with higher resolution of the TNG suite. It has previously been shown that galaxies in redshift space follow the so-called “negative binomial” hierarchical clustering model, but this agreement does not carry over to real space. This means that, under the VPF formulation, there is an apparent

scaling relation in redshift space between the mentioned two–point correlation function (the lowest clustering order) and the higher orders of clustering. Our results indicate, however, that galaxies within cosmic voids in real space follow the binomial model just like galaxies in redshift space; i.e., by studying galaxies in cosmic voids we recover the hierarchical scaling detected in redshift space.

As established, galaxies are of primary interest in the study of our Universe. Analysing their spatial and spectral light distributions in combination with their dynamical properties, one can infer the physical processes that govern galaxy formation and attempt to explain the plethora of galaxy types and structures one observes. Of particular interest are the orientations and ellipticities of galaxy images. Preferential orientations, or *alignments*, between galaxies, their underlying matter structures, and the aforementioned cosmic web are crucial aspects to further a more comprehensive understanding of gravity, the nature of matter, and structure formation in the Universe. We propose a set of well–behaved statistical parameters to study the alignment signal, with respect to the centers of voids, of spins of galaxies within shells around voids identified in the TNG300-1 simulation. We explore this signal in different samples of galaxies, varying their distance from the void center, mass, spin norm, local density, and velocity. We find a strong tendency ($> 9\sigma$) of massive, high–spin, and low radial velocity galaxies to be aligned perpendicularly to the void–centric direction when stacking shells of 0.9-1.4 void radii. Furthermore, we find that, in these subdense environments, local density is irrelevant to the alignment signal, while the largest impact on the alignment is due to radial velocity: galaxies that lose linear momentum with the void expansion seem to be strongly aligned perpendicularly to the center of the void. A further understanding of intrinsic alignments will be crucial for interpreting weak lensing studies and to capitalize on upcoming major cosmic shear surveys like Euclid or LSST.

Overall, we propose novel methods and statistics to study galaxy clustering on one hand, and orientations with respect to voids on the other, and finally utilize non–standard clustering measurements to characterize the clustering of galaxies inside voids compared to elsewhere in the cosmic web. We have learned that voids are a unique environment in which structure can form and evolve, and may play a crucial role in embedding galaxies with alignment and angular momentum.

*A mi padre Alejandro, a mi madre Mónica
A mis hermanos, Santiago y Gabriel, y a mi hermana Natalia*

Agradecimientos

Esta Tesis existe gracias a la democracia,
a la educación pública,
al apoyo incondicional de mis viejos,
al amor infinito de mi familia,
a la guía de mis directores Marce Lares y Diego García Lambas,
a la compañía de invaluable amigos y amigas y todo entremedio.

Gracias a todos ustedes,
y a la fatal y entrelazadora inevitabilidad de las coincidencias.

Índice general

Indice General	XIII
1. Del Universo homogéneo a las galaxias	1
1.1. Modelo cosmológico estándar	2
1.2. Dinámica de universos Friedman-Robertson-Walker	3
1.3. Parámetros Cosmológicos	6
1.4. El <i>redshift</i> cosmológico	7
1.5. Formación de estructuras	10
1.5.1. Campo de densidad	11
1.5.2. Evolución lineal de las perturbaciones	13
1.5.3. Evolución no lineal	14
1.5.4. Formación de galaxias	17
2. Patrones cósmicos, vacíos y galaxias	19
2.1. La Red Cósmica	19
2.2. Vacíos cósmicos	21
2.3. Alineaciones galácticas	24
2.3.1. Revisión de métodos y resultados	25
3. Estadísticos para el estudio de la distribución de materia	31
3.1. Función de correlación bipuntual (2PCF)	32
3.2. Función de correlación anisotrópica	33
3.2.1. 2PCF en el espacio de <i>redshift</i>	33
3.2.2. Multipolos de Legendre	38
3.3. Estadísticos de agrupamiento de alto orden	39
3.3.1. La función de probabilidad de vacío (VPF)	40
3.3.2. Escala jerárquica y la VPF reducida (RVPF)	41
3.3.3. Modelos fenomenológicos de agrupamiento	43

3.3.4.	Resultados previos	47
3.3.5.	Resumen y discusión	60
4.	Optimizando funciones de correlación: catálogos tipo <i>glass</i>	65
4.1.	Distribuciones puntuales homogéneas e isotrópicas	67
4.2.	Reconstrucción de Zeldovich	72
4.3.	Modificación del estimador de Landy-Szalay	74
4.4.	Rendimiento del estimador LS modificado	77
4.5.	Discusión	85
5.	Nuevos estadísticos para el estudio de alineamientos	88
5.1.	Metodología y nuevos estadísticos	89
5.1.1.	Cociente de los componentes vectoriales, β	91
5.1.2.	Cociente entre vectores perpendiculares y paralelos, η	96
5.1.3.	Una prueba para los coeficientes OLS de la distribución del coseno	104
5.2.	Aplicación a datos sintéticos, comparación con métodos tradicio- nales, y el parámetro ζ	109
5.3.	Discusión y comentarios	115
6.	Alineaciones de galaxias en voids	118
6.1.	Datos: TNG300-1	120
6.2.	Identificación de vacíos y su población de galaxias	120
6.2.1.	Velocidades y entorno de las galaxias	123
6.2.2.	Propiedades de las galaxias en los vacíos y su clasificación	123
6.3.	Efectos de alineaciones en las galaxias	124
6.3.1.	Alineación vs. masa	128
6.3.2.	Alineación vs. espín	128
6.3.3.	Alineación vs. densidad del entorno local	129
6.3.4.	Alineación vs. velocidad respecto al centro del vacío	129
6.3.5.	Alineación en combinaciones de muestras	130
6.4.	Discusión y comentarios	132
7.	Estructura con escalamiento jerárquico en vacíos cósmicos	138
7.1.	Metodología	139
7.1.1.	Pruebas de estabilidad de los estadísticos	140
7.1.2.	Periodicidad y simulación del espacio de <i>redshift</i>	142

7.2. Resultados	143
7.2.1. RVPF en el espacio real vs. en el espacio de <i>redshift</i>	144
7.2.2. RVPF dentro de vacíos cósmicos en el espacio real vs. en el espacio de <i>redshift</i>	149
7.2.3. Comparaciones finales	151
7.3. Conclusiones	156
8. Conclusiones	158
Bibliografía	164

Capítulo 1

Del Universo homogéneo a las galaxias

Motivaciones

Las galaxias que observamos en nuestro Universo son el fruto de una larga historia de procesos físicos, simples y complejos, que como tales, representan una puerta hacia entender no sólo la naturaleza de las galaxias, sino también la maquinaria menos visible de las leyes que las gobiernan.

En esta Tesis estudiamos propiedades de galaxias con un enfoque estadístico, desde la caracterización de su distribución en el espacio hasta sus propiedades astrofísicas, como el momento angular, y cómo éstas se relacionan con las estructuras circundantes. Además, esperamos aportar métodos estadísticos novedosos y robustos para estudios futuros de este tipo.

Previo a presentar las definiciones de los conceptos pertinentes, podemos esbozar de forma general las preguntas que abordamos con este trabajo:

- ¿Puede mejorarse el estudio de la distribución de las galaxias en el espacio y su orientación con respecto a estructuras circundantes?
- ¿Existen métodos alternativos para estos estudios? ¿Son igualmente válidos? ¿Qué información aportan y qué nos dicen acerca de las estructuras en vacíos cósmicos?
- ¿Están las galaxias orientadas aleatoriamente? ¿Cuál es su relación con estructuras en gran escala como los vacíos cósmicos?

A continuación haremos un breve repaso del contexto en el cual surgen estos interrogantes.

1.1. Modelo cosmológico estándar

La cosmología es la ciencia que estudia la composición, evolución y propiedades del Universo en escalas grandes. La principal hipótesis en esta ciencia es que las interacciones gravitatorias gobiernan la formación y evolución de estructuras en grandes escalas del Universo. Estas interacciones están descritas por la Teoría de la Relatividad de Einstein. La segunda hipótesis se denomina “principio cosmológico”, y asume que el universo es espacialmente homogéneo e isotrópico en grandes escalas. En particular, en esta Tesis hablamos sobre galaxias y las estructuras que forman en gran escala, es decir, trataremos con grandes fluctuaciones de densidades que parecen escapar a esta segunda hipótesis, por lo cual cabe remarcar que lo establecido por el principio cosmológico es que el Universo en grandes escalas es estadísticamente homogéneo e isotrópico.

Se conocen como “modelos cosmológicos” a los escenarios propuestos que explican y caracterizan distintos universos. En el caso de esta Tesis nos interesa un modelo cosmológico que de cuenta de la distribución observada de galaxias. El más aceptado actualmente es el modelo llamado *Lambda Cold Dark Matter* (Λ CDM). Este modelo propone una geometría espacial plana donde, conforme a la primera hipótesis, la evolución del espacio-tiempo queda determinada por la distribución de materia bajo el formalismo de la Relatividad General. La densidad total de materia incluye cuatro componentes: una densidad de radiación (ρ_γ), una densidad de materia bariónica (ρ_b), una densidad de materia oscura (ρ_{DM}) y una densidad de energía oscura (ρ_Λ). La densidad de radiación tuvo un rol protagónico en etapas tempranas del Universo pero para muchos estudios es despreciable en el presente. La materia bariónica está constituida por los protones, neutrones y electrones que conforman el Universo “visible” y conforma un $\sim 4\%$ de la composición total. La materia oscura ($\sim 21\%$) y la energía oscura ($\sim 75\%$), a pesar de ser los componentes dominantes en el Universo, son de naturaleza desconocida hasta el momento. La noción de la materia oscura (cuyo nombre señala su cualidad “invisible”) surge en la década de 1930 con el motivo de justificar la estabilidad gravitatoria observada en cúmulos de galaxias (Zwicky, 1933), las curvas de rotación de las galaxias (Babcock, 1939) y el acercamiento entre la Vía

Láctea y Andrómeda (Kahn y Woltjer, 1959). La noción de la energía oscura, por otro lado, fue presentada luego de encontrar evidencia de una expansión acelerada del Universo mediante estudios de la escala de distancias con supernovas (Perlmutter et al., 1999, Riess et al., 1998).

Como nos interesa contextualizar la distribución de galaxias observada, hacemos a continuación un breve repaso de cómo es la evolución dinámica de un universo conforme a este modelo cosmológico estándar (Sec. 1.2) para luego definir los parámetros que lo caracterizan (Sec. 1.3). Con los conceptos presentados en estas secciones podemos luego definir una cantidad crucial en la astronomía como es el *redshift* cosmológico (Sec. 1.4) y finalmente esquematizar la formación de estructuras y galaxias (Sec. 1.5).

1.2. Dinámica de universos Friedman-Robertson-Walker

Si conocemos el contenido de materia y energía del Universo, podemos determinar la evolución del mismo mediante las ecuaciones de campo de Einstein, que relacionan el contenido de energía y momento del espacio-tiempo con su geometría. Usando la convención $c = 1$ podemos escribir las ecuaciones de campo de la siguiente forma:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - g_{\mu\nu}\Lambda = 8\pi GT_{\mu\nu}, \quad (1.1)$$

donde $R_{\mu\nu}$ es el tensor de Ricci, que describe la geometría local del Universo, $g_{\mu\nu}$ es la métrica, R es el escalar de curvatura, $T_{\mu\nu}$ es el “tensor energía-momento” que abarca la información acerca de la energía y momento de la materia, y Λ es la denominada “constante cosmológica”, causante de la expansión acelerada observada en el Universo.

Para un universo homogéneo e isotrópico $g_{\mu\nu}$ está dado por la métrica de Friedmann-Robertson-Walker (FRW) (ver e.g. Weinberg, 1972). Con esta métrica el elemento de línea ds puede ser expresado como

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = dt^2 - a^2(t) \left[\frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \text{sen}^2(\theta)d\phi^2) \right], \quad (1.2)$$

donde $\{x^\mu\}_{\mu=0,1,2,3}$ son las coordenadas de observadores fundamentales en el espacio-tiempo ($\{t, r, \theta, \phi\}$ usando coordenadas esféricas), K es la constante que define si la curvatura del espacio-tiempo es abierta ($K = -1$), cerrada ($K = 1$) o plana ($K = 0$) y $a(t)$ es el denominado “factor de escala” que caracteriza la expansión del Universo. Estas dos cantidades determinan completamente la métrica de FRW.

Se pueden obtener ecuaciones para la evolución temporal de este factor de escala utilizando la métrica de FRW y el tensor de energía-momento de un fluido uniforme ideal

$$T_{\mu\nu} = \text{diag}(\rho, -p, -p, -p), \quad (1.3)$$

siendo ρ la densidad total de todos los componentes de la materia y p la presión del fluido. Con estas consideraciones, las ecuaciones de campo de Einstein se reducen a ecuaciones para $a(t)$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}, \quad (1.4)$$

$$H^2(t) \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2} + \frac{\Lambda}{3}, \quad (1.5)$$

donde el punto indica derivación respecto al tiempo cósmico t . Con la Ec. 1.5 se define el denominado “parámetro de Hubble”, $H(t)$. Finalmente, derivando esta ecuación y combinándola con 1.4 obtenemos la ecuación de conservación de la energía:

$$\dot{\rho} + 3H(\rho + p) = 0. \quad (1.6)$$

Para obtener soluciones explícitas para $a(t)$ y K podemos usar la Ec. 1.5, también conocida como la ecuación de Friedmann, y combinarla con la ecuación de la conservación de la energía, Ec. 1.6.

En esta instancia es necesario asumir alguna relación entre la presión y la densidad del fluido ideal (que satisface la Ec. 1.3) en la forma de una ecuación de estado $p = p(\rho)$. Con esta relación podemos integrar la Ec. 1.6 y obtener $\rho = \rho(p)$. En cosmología se suelen considerar fluidos en los que la presión es proporcional a la densidad, i.e. $p = \omega\rho$, llamados “fluidos barotrópicos” y que son caracterizados por tener una velocidad de sonido constante $c_s^2 = dp/d\rho$. La

cantidad ω que relaciona la presión con la densidad se denomina “parámetro de ecuación de estado”. Para estos fluidos, la conservación de la energía implica una relación entre la densidad y el factor de escala del tipo

$$\rho \propto a^{-3(1+\omega)}. \quad (1.7)$$

Este resultado es válido si asumimos N fluidos independientes con ecuaciones de estado ω_i . De esta forma, la densidad total de energía puede ser expresada como

$$\rho = \sum_{i=1}^N \frac{\rho_{i,0}}{a^{3(1+\omega_i)}}, \quad (1.8)$$

llamando $\rho_{i,0}$ a la densidad de cada fluido en el tiempo actual, t_0 y adoptando la convención de normalización del factor de escala $a_0 \equiv a(t_0) = 1$. Con esta expresión para la densidad de energía, la ecuación de Friedmann 1.5 puede escribirse como

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \sum_{i=1}^N \frac{\rho_{i,0}}{a^{3(1+\omega_i)}} - \frac{K}{a^2} + \frac{\Lambda}{3}, \quad (1.9)$$

Evaluando esta ecuación en el presente tenemos una expresión para el parámetro de Hubble en t_0 :

$$H_0^2 = \frac{8\pi G}{3} \sum_{i=1}^N \rho_{i,0} - K + \frac{\Lambda}{3}. \quad (1.10)$$

Este parámetro suele estar escrito en unidades de

$$H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}, \quad (1.11)$$

donde h es el denominado parámetro de Hubble adimensional y, según observaciones actuales, $h \simeq 0.7$ (ver, e.g., [Freedman y Madore, 2010](#), para una revisión de métodos de medición de h).

La Ec. 1.9 nos permite obtener una solución para K , es decir, una expresión para la curvatura del Universo en función del contenido de materia y energía, que, evaluada en el presente y asumiendo $\Lambda = 0$, se expresa como

$$K = H_0^2 \left(\frac{8\pi G}{3H_0^2} \sum_{i=1}^N \rho_{i,0} - 1 \right), \quad (1.12)$$

donde es fácil ver que la curvatura del espacio-tiempo está dada por el contenido energético del universo: la Ec. 1.12 describiría un universo plano, i.e. $K = 0$, siempre que la densidad total sea igual a un valor crítico, ρ_c , igual al inverso del multiplicador del primer término:

$$\begin{aligned} \rho_c &\equiv \frac{3H_0^2}{8\pi G} = 2.773 \times 10^{11} \text{M}_\odot h^{-2} \text{Mpc}^{-3} \\ &= 11.26 h^{-2} \text{protones } m^{-3}, \end{aligned} \quad (1.13)$$

donde como referencia, asumiendo $h \simeq 0.7$, podemos ver que la densidad crítica corresponde a un fluido con 6 protones por metro cúbico. Esta cantidad crítica ρ_c se utiliza para definir los parámetros cosmológicos, presentados en la subsección siguiente, que caracterizan un dado modelo.

1.3. Parámetros Cosmológicos

Los diferentes modelos cosmológicos están caracterizados por los parámetros definidos como el cociente de la densidad actual de cada componente del Universo y la densidad crítica ρ_c (Ec. 1.13). Es decir, los parámetros cosmológicos para la materia (componentes bariónicos y materia oscura), la radiación (bosones y partículas relativistas), la constante cosmológica y la curvatura están determinados respectivamente por las siguientes expresiones:

$$\Omega_m = \frac{8\pi G}{3} H_0^2 \rho_{m,0}, \quad (1.14)$$

$$\Omega_\gamma = \frac{8\pi G}{3} H_0^2 \rho_{\gamma,0}, \quad (1.15)$$

$$\Omega_\Lambda = \frac{\Lambda}{3H_0^2}, \quad (1.16)$$

$$\Omega_K = -\frac{K}{H_0^2}. \quad (1.17)$$

$$(1.18)$$

A partir de estas definiciones la ecuación de Friedmann puede reescribirse en términos de los parámetros cosmológicos y el factor de escala a como

$$H^2(a) = H_0^2[\Omega_m a^{-3} + \Omega_\gamma a^{-4} + \Omega_K a^{-2} + \Omega_\Lambda]. \quad (1.19)$$

Luego, una vez que la cosmología es especificada en términos de los parámetros Ω , la Ec. 1.19 puede integrarse para encontrar $a(t)$ y con ello determinar la evolución temporal del Universo. Además, con estas definiciones podemos escribir la Ec. 1.12 de una forma más compacta:

$$\sum_{i=1}^N \Omega_i + \Omega_\Lambda + \Omega_K = 1, \quad (1.20)$$

por lo tanto, aún para el caso $\Lambda \neq 0$, la curvatura será plana, es decir $\Omega_K = 0$, siempre que $\Omega_\gamma + \Omega_m + \Omega_\Lambda = 1$.

Para el modelo cosmológico estándar Λ CDM, que describe un Universo plano y en expansión compuesto por materia bariónica y materia oscura fría, se encuentra que los parámetros cosmológicos que describen el contenido del Universo tienen los valores expresados a continuación: $\Omega_m = 0.3089 \pm 0.0062$, $\Omega_\Lambda = 0.6911 \pm 0.062$, $H = 67,74 \pm 0,46 \text{ km s}^{-1} \text{ Mpc}^{-1}$ de acuerdo a las mediciones del Fondo Cósmico de Microondas (CMB por su nombre en inglés, *Cosmic Microwave Background*) (Planck Collaboration et al., 2016).

1.4. El *redshift* cosmológico

El factor de escala $a(t)$ es de crucial utilidad a la hora de medir los cambios en frecuencia de la luz emitida por fuentes distantes. Hubble detectó, no sólo que las líneas espectrales de galaxias cercanas a la Vía Láctea se desplazaban a longitudes de onda mayores, sino que lo hacían proporcionalmente a la distancia a la que se encontraban, por lo que no podía explicarse sólo mediante el efecto Doppler (Hubble, 1929). Este descubrimiento demostró empíricamente la expansión del Universo, y que es esta expansión la causante del enrojecimiento de las líneas espectrales.

Llamaremos de ahora en adelante *redshift* (palabra en inglés refiriéndose a un desplazamiento al rojo) a este aumento en las longitudes de onda, o disminución en la frecuencia, efectivamente enrojeciendo la señal emitida desde la fuente, denotándolo con la letra z de la siguiente forma:

$$z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{\nu_1 - \nu_0}{\nu_0}. \quad (1.21)$$

Para calcular este cambio en frecuencia consideremos la ecuación de movimiento (dada por la métrica FRW, Ec. 1.2) para una dada cresta de una onda electromagnética viajando hacia el observador situado en un origen $\mathbf{r} = 0$, teniendo en cuenta que para un fotón se cumple la igualdad $ds^2 = 0$:

$$0 = cdt - a^2(t) \frac{dr^2}{1 - Kr^2}. \quad (1.22)$$

Luego, si llamamos r_1 a la ubicación de la fuente respecto al observador y t_1 al momento de emisión, la onda llegará a nosotros a un tiempo t_0 dado por

$$\int_{t_1}^{t_0} \frac{cdt}{a(t)} = f(r_1) \quad (1.23)$$

donde

$$f(r_1) = \int_0^{r_1} \frac{dr}{\sqrt{1 - Kr^2}} = \begin{cases} \arcsen(r_1), & \text{si } K = +1 \\ r_1, & \text{si } K = 0 \\ \operatorname{arcsenh}(r_1), & \text{si } K = -1 \end{cases} \quad (1.24)$$

Notamos que $f(r_1)$ es independiente del tiempo ya que la dependencia temporal está contenida en el factor de escala.

Consideremos ahora una segunda cresta de la onda electromagnética emitida por la fuente en la posición r_1 en un tiempo $t_1 + \delta t_1$. Esta llegará al observador al tiempo $t_0 + \delta t_0$ dado por

$$\int_{t_1 + \delta t_1}^{t_0 + \delta t_0} \frac{cdt}{a(t)} = f(r_1). \quad (1.25)$$

Si asumimos un intervalo temporal lo suficientemente chico tal que los cambios en $a(t)$ sean despreciables y restamos las Ecs. 1.23 y 1.25 obtenemos:

$$\frac{c\delta t_0}{a(t_0)} = \frac{c\delta t_1}{a(t_1)}. \quad (1.26)$$

De esta expresión se desprende que el cociente entre la frecuencia observada, ν_0 , y la emitida, ν_1 , es

$$\frac{\nu_0}{\nu_1} = \frac{\delta t_1}{\delta t_0} = \frac{a(t_1)}{a(t_0)}. \quad (1.27)$$

En un universo en expansión se cumple la desigualdad $a(t_1) > a(t_0)$, lo que implica que $\nu_1 < \nu_0$, es decir, la frecuencia de la onda recibida se encuentra enrojecida respecto a su frecuencia original.

Con esta relación y la definición de *redshift* de la Ec. 1.21, es fácil ver la dependencia de este con el factor de escala en los tiempos t_0 y t_1 :

$$z = \frac{a(t_0)}{a(t_1)} - 1, \quad (1.28)$$

y si adoptamos t_0 como el tiempo presente y recordamos la convención adoptada $a(t_0) = 1$, obtenemos la relación entre la expansión del Universo y el enrojecimiento que esta causa en las líneas espectrales de las galaxias,

$$z = \frac{1}{a(t)} - 1. \quad (1.29)$$

Esta expresión nos indica que la luz que recibimos hoy de una galaxia con *redshift* z fue emitida cuando el factor de escala universo era un factor $1/(1+z)$ de su valor actual.

Dado que el *redshift* es un observable directo, en muchos casos es más conveniente usarlo como variable en lugar de t . Con la relación

$$H(z) = -\frac{1}{(1+z)} \frac{dz}{dt}, \quad (1.30)$$

y la Ec. 1.23, tenemos la relación entre la distancia a un objeto y su *redshift*:

$$D(z) = \int_0^z \frac{cdz'}{H(z')}, \quad (1.31)$$

con

$$H(z) = H_0 \left[\sum_{i=0}^N \Omega_i (1+z)^{3(1-\omega_i)} + (\Omega_K + \Omega_\Lambda)(1+z)^2 \right]^{\frac{1}{2}}. \quad (1.32)$$

Para *redshifts* pequeños, podemos aproximar la distancia por

$$D(z) \simeq zD_H, \quad (1.33)$$

donde D_H es la “distancia de Hubble” dada por

$$D_H \equiv \frac{c}{H_0}. \quad (1.34)$$

Con la relación entre el factor de escala y el *redshift* podemos también determinar el tiempo transcurrido desde la emisión de la onda hasta ahora:

$$t(z) = \int_z^\infty \frac{dz'}{(1+z')H(z')}, \quad (1.35)$$

lo que también podría entenderse como la edad del Universo a la que se observa la fuente (Hogg, 1999).

1.5. Formación de estructuras

Con esta sección finalizamos el recorrido esquemático desde el universo homogéneo a la formación de estructuras y galaxias que estudiamos en esta Tesis.

Es ampliamente aceptado que en los orígenes del Universo existieron pequeñas fluctuaciones en la densidad que, mediante el proceso de inestabilidad gravitacional, crecieron hasta formar la estructuras que hoy observamos. Actualmente, el mecanismo que mejor explica los orígenes y evolución de estas fluctuaciones primordiales en el Universo es el mecanismo de inflación (ver e.g. Riotto, 2002, Tsujikawa, 2003, para una revisión de la teoría inflacionaria). La evidencia observacional clave que confirmó la idea de un universo inicialmente cuasi homogéneo es el Fondo Cósmico de Microondas (CMB), descubierto de forma accidental por los radioastrónomos Arno Penzias y Robert Wilson (Penzias y Wilson, 1965) confirmando la predicción teórica de la existencia de dicha radiación (Dicke et al., 1965). Este fondo de radiación cósmica nos permite ver cómo era el Universo cuando tenía ~ 300000 años de vida ($z \sim 1100$), mostrando que las fluctuaciones primordiales en temperatura eran del orden de $\sim 10^{-5}$.

Recordemos que los principios de homogeneidad e isotropía del Universo en las que se basa el modelo cosmológico estándar son sólo válidas en escalas de cientos de megaparsecs. En escalas menores encontramos una amplia variedad de estructuras con un gran rango de masas y tamaños. En otras palabras, la distribución de galaxias en estas escalas no es uniforme, sino que presenta una amplia variedad de estructuras dentro de las cuales se distinguen cúmulos, filamentos y grandes regiones subdensas. Sin embargo, el CMB nos muestra un escenario en el que la distribución de materia y energía del Universo primitivo presenta sólo

fluctuaciones muy pequeñas. De esto se deduce que la amplia variedad de estructuras que observamos en la distribución de galaxias es indicativo de la existencia de procesos físicos que llevaron al Universo desde un estado inicial casi totalmente homogéneo a una compleja red de estructuras.

Dentro del paradigma establecido por el modelo cosmológico Λ CDM, la teoría predice un proceso de formación jerárquica en donde las estructuras pequeñas son las primeras en desacoplarse de la expansión de Hubble para luego colapsar y alcanzar un equilibrio dinámico cercano al equilibrio virial. Este proceso implica un escenario jerárquico de formación en el que las estructuras de mayor tamaño se forman a partir de agrupaciones de estructuras más pequeñas ya virializadas mediante acreción de materia.

Para describir el proceso de formación de estructuras siguiendo el crecimiento de fluctuaciones primordiales pequeñas, resulta conveniente utilizar el sistema de coordenadas “comóvil”, i.e. que sigue la expansión del Universo. Definimos las posiciones \mathbf{x} como

$$\mathbf{x} = \frac{\mathbf{r}}{a(t)}, \quad (1.36)$$

donde \mathbf{r} es la posición en un sistema de coordenadas estático y $a(t)$ es el factor de escala introducido en la Ec. 1.2.

Al trabajar en un sistema de coordenadas comóvil podemos desglosar la velocidad \mathbf{v} en una componente que es producto de la expansión uniforme de Hubble y en otra componente peculiar, i.e. propia de la galaxia, \mathbf{u} , que describe el apartamiento de esta expansión isotrópica:

$$\mathbf{v} = \frac{\dot{a}}{a}\mathbf{r} + a(t)\dot{\mathbf{x}} = H(t)\mathbf{r} + \mathbf{u}. \quad (1.37)$$

1.5.1. Campo de densidad

Definimos las fluctuaciones de densidad respecto a la densidad media de fondo como:

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t)}{\bar{\rho}(t)} - 1 \quad (1.38)$$

donde $\bar{\rho}(t)$ es la densidad media del Universo y $\rho(\mathbf{x}, t)$ es la densidad en la posición \mathbf{x} , ambas en un mismo tiempo t . Notamos que, según esta definición, el

contraste de densidad toma valores positivos para sobredensidades y negativos para subdensidades.

Las fluctuaciones pueden describirse, según la teoría inflacionaria, por un campo gaussiano aleatorio. Es decir, por un campo en el que los diferentes modos de Fourier son estadísticamente independientes, cada uno con una fase aleatoria. Con estas suposiciones, las propiedades estadísticas del campo de densidad primordial quedan descritas completamente por el espectro de potencia $P(k)$ definido como:

$$P(k) \equiv \langle |\delta(\mathbf{k})|^2 \rangle, \quad (1.39)$$

donde \mathbf{k} denota un vector de onda cuyo módulo, el número de onda $k = |\mathbf{k}|$, caracteriza una dada escala en el espacio, y

$$\delta(\mathbf{k}) = \int_{-\infty}^{\infty} d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} \delta(\mathbf{x}). \quad (1.40)$$

Podemos ver en esta ecuación que el espectro de potencias $P(k)$ describe la amplitud de las fluctuaciones del campo de densidad en las escalas $\lambda = 2\pi/k$.

El espectro de potencia inicial, $P(k) \propto k^{n_s}$ donde el “índice espectral” n_s describe cómo varían las fluctuaciones de densidad con la escala, depende de cómo fueron generadas las fluctuaciones iniciales. El mecanismo estándar de inflación produce un espectro de potencia invariante respecto a la escala, con $n_s \sim 1$, llevándonos al $P(k)$ inicial conocido como espectro de Harrison-Zeldovich. El espectro de potencia inicial cambia a medida que el Universo se expande y las perturbaciones crecen. La geometría del Universo, y la materia y la energía en él suprimen el crecimiento de las perturbaciones en algunas escalas particulares y esto puede representarse con la expresión

$$P(k, t) = P_i(k) T^2(k, t), \quad (1.41)$$

donde se denomina $T(k, t)$ a la Función de Transferencia Lineal que relaciona el espectro de potencia $P(k, t)$ en un dado tiempo t con su forma inicial $P_i(k)$.

Los modelos actuales de la inflación no predicen una amplitud específica del espectro de potencia primordial. Se establece de forma observacional mediante las fluctuaciones de la temperatura del CMB o a partir de mediciones locales de la estructura en gran escala. En este último caso, la amplitud del espectro de potencias $P(k)$ es determinada a partir de la varianza de la masa o fluctuación cuadrática media del campo de densidad en la escala $R_* = 8h^{-1}\text{Mpc}$ utilizando

una función ventana W como filtro para suavizar el campo de fluctuaciones $\delta(\mathbf{x})$. Este valor, denominado σ_8 , al igual que n_s es un valor cosmológico fundamental que permite determinar la evolución de las fluctuaciones primordiales:

$$\sigma_8^2 \equiv \left\langle \left(\frac{\delta\rho}{\rho} \right)^2 \right\rangle = \frac{1}{(4\pi)^2} \int_0^\infty dk W^2(kR_*) P(k) k^2. \quad (1.42)$$

1.5.2. Evolución lineal de las perturbaciones

Suponiendo que la materia puede aproximarse por un fluido ideal, no-colisional y autogravitante, y además que las perturbaciones son pequeñas y de escalas mucho menores al radio de Hubble (cH^{-1}), podemos realizar un tratamiento newtoniano de la evolución de las perturbaciones como caso límite de la Teoría de la Relatividad General (Peebles, 1980). De esta forma, la evolución de las fluctuaciones primordiales está descripta por las ecuaciones:

$$\frac{\partial\delta}{\partial t} + \frac{1}{a} \nabla \cdot [(1 + \delta)\mathbf{u}] = 0 \quad (1.43)$$

$$\frac{\partial\mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}(\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{1}{a}\nabla\phi \quad (1.44)$$

$$\nabla^2\phi = 4\pi G\bar{\rho}a^2\delta. \quad (1.45)$$

Donde \mathbf{u} es la velocidad peculiar de la partícula como en la Ec. 1.37. La primera de estas ecuaciones es la ecuación de continuidad que expresa la conservación de masa (Ec. 1.43). La segunda es la ecuación de movimiento para el fluido conocida como ecuación de Euler (Ec. 1.44). La tercera es la ecuación de Poisson (Ec. 1.45), que obedece el potencial gravitacional ϕ .

La validez de esta descripción se mantiene hasta el momento en que se produce el cruce de las trayectorias de las partículas o “cruce de cáscaras”. Este cruce sucede generalmente una vez que las perturbaciones crecieron lo suficiente para pasar al régimen no lineal. A partir de este momento debe seguirse la trayectoria de cada partícula individualmente.

Sin embargo, si consideramos $\delta \ll 1$, es decir, desviaciones pequeñas de la homogeneidad e isotropía, podemos linealizar las Ecs. 1.43 y 1.44 considerando δ y \mathbf{u} sólo hasta primer orden y combinarlas para obtener una única ecuación que describe la evolución de las perturbaciones con el tiempo:

$$\frac{\partial^2 \delta}{\partial t^2} + 2H \frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho} \delta. \quad (1.46)$$

La linealidad de esta ecuación en las perturbaciones δ implica que es separable y la solución general puede ser expresada como:

$$\delta(\mathbf{x}, t) = f_1(\mathbf{x})D_1(t) + f_2(\mathbf{x})D_2(t), \quad (1.47)$$

donde $D_1(t)$ y $D_2(t)$ son soluciones independientes de la Ec. 1.46, una creciente y otra decreciente. Debido a que los efectos gravitatorios hacen crecer a las sobredensidades, el modo decreciente tiende a cero con el tiempo. Por lo tanto, el crecimiento de las perturbaciones escala con el tiempo en el régimen lineal como:

$$\delta(t) = \delta(t_0) \frac{D_1(t)}{D_1(t_0)}, \quad (1.48)$$

donde D_1 es el modo creciente y t_0 es el tiempo en el presente. Es decir, durante el crecimiento lineal las perturbaciones de densidad mantienen su forma en las coordenadas comóviles y crecen proporcionalmente a $D_1(t)$.

La función $D_1(t)$, llamada factor de crecimiento, describe la evolución de las fluctuaciones en el Universo temprano y en grandes escalas, donde la aproximación $\delta \ll 1$ también es válida (ver Hamilton, 2001, para una síntesis de expresiones del factor de crecimiento para distintas relaciones entre parámetros cosmológicos).

1.5.3. Evolución no lineal

En la subsección anterior resumimos un desarrollo analítico de la evolución de las perturbaciones del campo de densidad dentro del régimen lineal, i.e. $\delta \ll 1$. Este desarrollo lineal pierde validez en escenarios donde los contrastes de densidad son del orden de la unidad, instancia en la cual debemos acudir a otros métodos para analizar la evolución de las perturbaciones. Generalmente se aborda el problema mediante grandes simulaciones numéricas (e.g. Hamilton et al., 1991, Moore et al., 1999, Navarro et al., 1996). Sin embargo, hay algunas aproximaciones analíticas útiles para numerosos casos. A continuación describiremos la aproximación de Zeldovich (Zel'dovich, 1970), que provee una perspectiva general de la formación de estructuras en el Universo. En el Cap. 4 presentamos un proceso denominado “Reconstrucción de Zeldovich” que usa esta aproximación y lo implementamos en un método para disminuir la varianza de estadísticos de

agrupación de galaxias.

Aproximación de Zeldovich

La aproximación de Zeldovich es una descripción simple pero efectiva de la evolución de las perturbaciones del campo de densidad que consiste en una extrapolación de la inestabilidad gravitacional lineal hacia el régimen no lineal siguiendo la trayectoria de las partículas que componen el fluido del campo continuo de densidad. Presentada por Zel'dovich (1970), esta aproximación cinemática resuelve las ecuaciones de movimiento de una distribución de materia inicialmente homogénea y no colisional empleando una formulación lagrangiana de las perturbaciones a primer orden.

La aproximación supone que las posiciones comóviles iniciales de las partículas $\mathbf{q}(t = 0)$ son desplazadas a las coordenadas eulerianas en el tiempo $\mathbf{r}(t)$ dadas por:

$$\mathbf{r}(t) = a(t)[\mathbf{q} + b(t)\mathbf{f}(\mathbf{q})], \quad (1.49)$$

donde, por construcción, \mathbf{q} se mantiene constante a través del tiempo. Esta ecuación describe la expansión de Hubble con una perturbación dada por el segundo término: $b(t)\mathbf{f}(\mathbf{q})$. Esta función $b(t)$, despreciable para $t \rightarrow 0$, escala el campo de desplazamiento $\mathbf{f}(\mathbf{q})$. Este campo, a su vez, es independiente del tiempo y se relaciona con el potencial de desplazamiento originado por las fluctuaciones iniciales mediante $\mathbf{f}(\mathbf{q}) = -\nabla\Phi(\mathbf{q})$.

La Ec. 1.49 representa un mapeo entre las coordenadas eulerianas y lagrangianas que es único siempre y cuando las trayectorias no se crucen, en cuyo caso pueden producirse campos multivariados y singularidades en el campo de densidad. Esto está predicho por la aproximación y puede estimarse mediante la conservación de masa. Para ello consideremos un elemento de masa en un dado volumen tal que la densidad $\delta(\mathbf{r}, t)$ sea igual a la densidad media inicial del Universo $\bar{\rho}(0)$, debido a la conservación de la masa tenemos entonces:

$$\rho(\mathbf{r}, t) = \bar{\rho}(0)d\mathbf{q}. \quad (1.50)$$

Luego, expresando el campo de densidad en función de las coordenadas lagrangianas, tenemos:

$$\rho(\mathbf{q}, t) = \bar{\rho}(0) J \left[\frac{\partial \mathbf{r}}{\partial \mathbf{q}} \right]^{-1}, \quad (1.51)$$

donde $J|\partial \mathbf{r}/\partial \mathbf{q}|$ es el determinante del jacobiano de la transformación entre las coordenadas eulerianas y lagrangianas, que tiene en cuenta la evolución gravitacional del fluido. Así,

$$\rho(\mathbf{q}, t) = \frac{\bar{\rho}}{|\delta_{ij} - b(t)\partial \mathbf{f}_i/\partial \mathbf{q}_j|}, \quad (1.52)$$

donde δ_{ij} la función Delta de Kronecker y $\bar{\rho} = (a(0)/a)^3 \rho(0)$ es la densidad media en el tiempo t .

La Ec. 1.52 nos dice que en las regiones donde se cumpla $\partial \mathbf{f}_i/\partial \mathbf{q}_j < 0$ el contraste de densidad con respecto al fondo aumentará, y no hay una cota superior para este aumento. Generalmente los términos $\partial \mathbf{f}_i/\partial \mathbf{q}_j$ son diferentes y el colapso comenzará a lo largo del eje definido por el término mayor. Como consecuencia, en la aproximación de Zeldovich la evolución gravitatoria de la perturbación propicia la formación de estructuras aplanadas conocidas como los “panqueques de Zeldovich”. Si el colapso se produce a lo largo de dos ejes, la evolución de las perturbaciones da lugar a la formación de filamentos, usualmente en la intersección de dos panqueques. Por último, el colapso puede suceder a lo largo de las tres direcciones simultáneamente, generando estructuras que se encuentran localizadas en la intersección de dos o más filamentos.

La aproximación de Zeldovich es efectiva para predecir el campo de densidad especialmente hasta la formación de los denominados panqueques y las otras estructuras mencionadas. Sin embargo, al ser un método puramente cinemático, las partículas se mueven en la misma dirección que en la condición inicial haciendo que las estructuras existan de manera instantánea para luego desarmarse. Dado su buen comportamiento hasta el cruce de cáscaras, este método ha sido ampliamente utilizado para generar condiciones iniciales de simulaciones numéricas de N-cuerpos. Sin embargo, el hecho de que sea un tratamiento puramente cinemático implica que las partículas se moverán siempre en la dirección determinada por la condición inicial. Esto significa que las estructuras existen sólo momentáneamente en el instante de su formación. Aún así, el buen comportamiento de esta aproximación hasta el cruce de cáscaras llevó a que sea usado comúnmente para la generación de condiciones iniciales en simulaciones numéricas de N-cuerpos. En el Cap. 4 la empleamos para reducir ruido poissoniano en distribuciones alea-

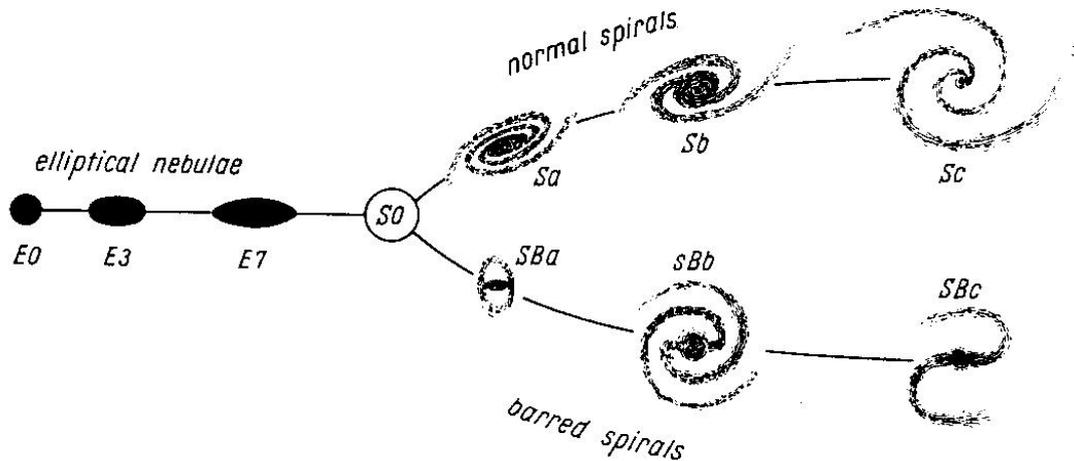


Figura 1.1: Clasificación morfológica de Hubble (llamada también “secuencia de Hubble”, de Hubble (1936). Esta clasificación divide las galaxias, según su forma, en elípticas (E), espirales (S) y espirales con barra (SB). Las elípticas se clasifican, según su elipticidad, de 0 a 7 en aumento de la misma. Las espirales y espirales barradas se subdividen en tres categorías: las Sa y las SBa presentan una predominancia del bulbo o núcleo respecto a los brazos, mientras que en las Sc y SBc predomina el núcleo. Las clasificación S0 corresponde a aquellas galaxias que poseen un disco pero no presentan brazos como las galaxias espirales. Existen en la literatura numerosas clasificaciones intermedias y, en particular, una categoría no mostrada en la figura que es la de galaxias irregulares. Esta clasificación corresponde a galaxias que no presentan de forma clara ninguna de las mencionadas características morfológicas.

torias, efectivamente homogeneizándolas, para reducir la varianza o el tiempo de cómputo de estadísticos de agrupamiento como la función de correlación.

1.5.4. Formación de galaxias

Parte del legado más ampliamente conocido y usado que dejó Hubble consistió en una clasificación morfológica de galaxias, mostrada en la Fig. 1.1, producto de sus estudios sistemáticos de las mismas (Hubble, 1926). Si bien durante muchos años se la interpretó como una clasificación estática, actualmente sabemos que las galaxias pueden cambiar sus características morfológicas a lo largo de su vida dependiendo del entorno donde se forman y evolucionan.

Los bariones reaccionan gravitatoriamente de la misma manera que la materia oscura, por lo tanto es de esperar que sigan su distribución en gran escala y que las galaxias se formen en los centros de halos formados por materia oscura. Se define como halo de materia oscura a las regiones donde se ha concentrado una

gran cantidad de materia gravitacionalmente ligada, sirviendo como trazadores (sesgadas) del campo de densidad y formando las unidades no-lineales fundamentales que sirven de sustento al escenario actual de formación de galaxias (White y Rees, 1978). Considerando esto, podemos encontrar fundamentos a la clasificación de Hubble. A grandes rasgos, las galaxias espirales se forman en halos que se encuentran relativamente aislados y tienen una historia de formación con poca actividad, de tal manera que el gas puede condensarse de manera suave y formar discos, mientras que las galaxias elípticas se formarían en regiones más pobladas donde son frecuentes las interacciones y fusiones con otros halos, transfiriendo energía y momento angular a la materia bariónica y oscura permitiendo así que los remanentes de estrellas y gas formen objetos más compactos y con baja rotación (Fall y Efstathiou, 1980).

En la Sec. 1.1 mencionamos que el $\sim 21\%$ del Universo conocido está compuesto por materia oscura y sólo $\sim 4\%$ por materia bariónica. Esto significa que los bariones constituyen solo el $\sim 16\%$ del contenido total de materia, por lo que son una minoría cuando se analiza la distribución media de materia en halos. Sin embargo, cerca de los centros de los halos encontramos una situación inversa: la región visible de las galaxias brillantes parece estar dominada por bariones (Mo et al., 2006). En el colapso del halo, los bariones asociados a él se compactan, enfrían y condensan en el centro para formar las galaxias; esta condensación central del gas causa una compresión adiabática en las partes internas del halo de materia oscura de tal manera que la densidad media dentro de la galaxia sería substancialmente mayor que en ausencia de bariones (Blumenthal et al., 1986, Mo et al., 1998). Los modelos de formación y evolución de galaxias basados en simulaciones de N-cuerpos deberían incluir los procesos, inducidos por bariones, de compresión adiabática, perturbación de marea y reionización que impactan la abundancia y estructura interna de subhalos de materia oscura (Zhu et al., 2016).

Capítulo 2

Patrones cósmicos, vacíos y galaxias

2.1. La Red Cósmica

La disposición espacial de la masa y las galaxias en configuraciones diversas tales como filamentos alargados, paredes laminares y cúmulos densos y compactos, la existencia de grandes regiones subdensas y la naturaleza jerárquica de esta distribución marcada por subestructuras abarcando una amplia gama de escalas y densidades, son tres características principales de lo que se ha denominado la Red Cósmica (Bond et al., 1996, van de Weygaert y Schaap, 2009).

Esta enorme Red Cósmica es definitivamente uno de los ejemplos más interesantes y llamativos de patrones cósmicos que se han encontrado en la naturaleza. La organización y distribución espacial de la materia en cualquier sistema físico es una exteriorización directamente observable que codifica información de la combinación de procesos y fuerzas que la formaron. Esto ha llevado a que los patrones geométricos se conviertan, en muchos y diversos estudios de índole científica, en una veta de crucial importancia para explorar y descubrir la física subyacente (e.g. Balbus y Hawley, 1998, van de Weygaert y Schaap, 2009).

El universo homogéneo, isotrópico y con perturbaciones de densidades propias del régimen lineal, esbozado en el Cap. 1, ha sido comprendido con suficiente profundidad como para extraer parámetros cosmológicos con increíble precisión (e.g. Planck Collaboration et al., 2016). Por otra parte, el universo inhomogéneo y anisotrópico de las estructuras filamentosas, laminares y grandes depresiones de densidad conforman un escenario de inestabilidad gravitatoria ligeramente avan-

zada, y la descripción no lineal de la formación y evolución de estas estructuras tiene considerables dificultades numéricas. No obstante, un caudal de información valiosa puede encontrarse explorando esta etapa levemente no lineal de la evolución de la estructura en la cual empiezan a emerger características individualmente distinguibles (van de Weygaert y Bond, 2008).

La existencia de los llamativos patrones y entramados representativos de esta etapa temprana de la evolución no lineal fue puesta de manifiesto gracias al empeño de crear mapas de la distribución de galaxias en gran escala (e.g. de Lapparent et al., 1986, Geller y Huchra, 1989, Huchra, 1989, Huchra et al., 1983, Huchra et al., 1994, York et al., 2000, entre otros). A su vez, las simulaciones cada vez más grandes de N -cuerpos mostraron que tales distribuciones son efectivamente manifestaciones de la evolución mediante inestabilidad y agrupación gravitacional (e.g. Dave et al., 2019, Dubois et al., 2014, Nelson et al., 2018, y más).

No obstante estos grandes avances, una comprensión física y una visión completa siguen justo fuera de nuestro alcance. Inconvenientes como la dificultad para establecer simetrías e influencias fuertes no locales suponen un gran obstáculo para el desarrollo de descripciones analíticas relevantes (van de Weygaert y Schaap, 2009). Además, la naturaleza jerárquica del proceso de agrupación gravitacional implica que no podemos considerar sólo una escala característica sino que hay formación de distintas estructuras en distintas escalas que hay que tener en cuenta. Existen aproximaciones analíticas que permiten tener una visión general acerca de las interacciones complejas entre las estructuras emergentes en todo el Universo en distintas escalas, así como también simulaciones que nos proveen una buena idea acerca de las complejidades de dicha distribución de materia emergente. Sin embargo, para el análisis de los patrones que resultan de estas interacciones complejas y subestructuras jerárquicas, la descripción sigue siendo en gran parte de naturaleza heurística (van de Weygaert y Schaap, 2009).

A diferencia de la cosmología que emplea campos continuos para describir el desarrollo de estructuras, el estudio de la estructura del Universo en gran escala suele basarse principalmente en muestreos discretos de estos campos de densidad. En particular, las galaxias suelen ser las trazadoras principales del gran entramado cósmico, y las estructuras que conforman han sido medidas a través de la distribución de sus *redshift*. Asimismo, las simulaciones de la evolución de la distribución de materia están basadas en gran medida en un cálculo computacional

de N-cuerpos, lo cual involucra una representación discreta de las características de interés para los estudios de la estructura en gran escala (van de Weygaert y Schaap, 2009).

Esta clase de estudios observacionales y simulaciones computacionales de N-cuerpos han puesto en evidencia que las galaxias se encuentran en grupos y cúmulos compactos, filamentos, y paredes laminares que rodean grandes regiones subdensas llamados vacíos cósmicos. En particular puede verse un ejemplo de los rasgos característicos de la Red Cósmica en la Fig. 7.2 que muestra una rodaja de la simulación TNG300-1 (Nelson et al., 2018) que describiremos más exhaustivamente en capítulos posteriores. Para el estudio de la formación de estructura en el Universo es de suma importancia el hecho de que los constituyentes estructurales de la Red Cósmica no se dispersan de forma aleatoria e independiente, sino que conforman estos patrones reconocibles a lo largo y ancho de todo el Universo conocido.

Las enormes regiones subdensas que dominan en volumen a la distribución de galaxias representan componentes que ofrecen información complementaria y contrastan las estructuras laminares o planares y filamentarias más densas (van de Weygaert y Schaap, 2009). Centramos nuestro enfoque ahora en ellos.

2.2. Vacíos cósmicos

En temas de estructura del Universo en gran escala, se llama por simplicidad “vacíos”, o por su nombre en inglés *voids*, a los mencionados vacíos cósmicos. Adoptamos esta convención para gran parte de esta Tesis, con algunas excepciones (ver Sec. 3.3) que serán oportunamente explicitadas.

Los vacíos pueden ser descriptos, en una muy primera aproximación, como regiones subdensas con simetría esférica y diámetros que varían entre algunos pocos Mpc a cientos de Mpc dependiendo del método de identificación y la densidad de la muestra. Rodeados por paredes, filamentos y cúmulos de galaxias, los vacíos son una de las componentes dominantes de la estructura en gran escala.

Los primeros análisis estadísticos se realizaron una vez que los catálogos de galaxias fueron suficientemente grandes y detallados, ampliando las muestras de vacíos (Hoyle et al., 2012, Hoyle y Vogeley, 2004, Pan et al., 2012). Desde entonces el estudio de los vacíos ha estado sujeto a una amplia variedad de definiciones donde éstos pueden ser esféricos, pudiendo o no superponerse entre sí, o tener

formas menos regulares. Algunos autores (ver, e.g, Colberg et al., 2008), además, imponen ciertas restricciones a la presencia de galaxias o a las propiedades de las galaxias en los vacíos, como por ejemplo la luminosidad o densidad local. Sin embargo, pueden sacarse características en común de las distintas definiciones (Ceccarelli et al., 2013, Colberg et al., 2008):

- Los vacíos son regiones de extrema subdensidad donde el contraste de densidad toma valores $\delta \lesssim -0.9$.
- Los valores mínimos del contraste de densidad conforman las zonas centrales del vacío.
- La dinámica de los vacíos queda determinada por la expansión de la subdensidad.

De forma análoga al caso de la distribución de galaxias, grandes vacíos pueden encontrarse en la distribución de materia y halos de materia oscura de las simulaciones numéricas (Aragon-Calvo et al., 2010, Bertschinger, 1985, Ceccarelli et al., 2006, Hoyle y Vogelely, 2004, Sheth et al., 2004). Teniendo en cuenta que la distribución de galaxias traza, de manera sesgada, la de los halos de materia oscura, Padilla et al. (2005) estudiaron y compararon las principales características dinámicas y espaciales de los vacíos en la distribución de los halos de materia oscura. Encuentran que esta distribución produce muestras de vacíos similares a las identificadas en la distribución de galaxias.

Identificación de vacíos

La identificación de vacíos ya sea en la distribución de galaxias o en la de materia abarca un gran abanico de métodos de búsqueda de regiones subdensas empleando algún criterio de densidad, geometría o dinámica. Ver Colberg et al. (2008) para una síntesis comparativa de los algoritmos más populares de identificación de vacíos en el espacio real. Muchos de estos algoritmos identifican regiones vacías de galaxias (El-Ad y Piran, 1997, Hoyle y Vogelely, 2002, Kauffmann y Fairall, 1991) o identifican determinadas estructuras geométricas presentes en la distribución de materia mediante el uso de galaxias como trazadoras de la misma (Colberg et al., 2005, Neyrinck, 2008, Plionis y Basilakos, 2002, Shandarin et al., 2006).

En esta Tesis hemos empleado el algoritmo de Ruiz et al. (2015) que es una versión modificada del procedimiento presentado por Padilla et al. (2005) y Ceccarelli et al. (2006) que consiste en buscar e identificar regiones esféricas subdensas sin solapamiento entre ellas y con un contraste de densidad integrado $\Delta = -0.9$. Describimos en detalle este algoritmo en la Sec. 6.2 donde lo aplicamos a la simulación TNG300-1 (ver Sec. 6.1) y desarrollamos el concepto de clasificación de vacíos tipo R y S .

Formación de vacíos

De lo resumido en la Sec. 1.5 y sus subsecciones, podemos decir que, en el modelo estándar, si las sobredensidades emergen de pequeñas fluctuaciones positivas en el campo primordial de densidad, entonces los vacíos emergen de pequeñas fluctuaciones que toman valores negativos. La gravedad se interpreta siempre como una “fuerza” de carácter atractivo. Sin embargo, gracias al déficit en densidad respecto a la distribución de materia de fondo, cuando tratamos con estas regiones subdensas se habla de una influencia gravitacional *efectiva* de carácter expansivo. Como consecuencia, con el crecimiento de las fluctuaciones por inestabilidad gravitacional se producen regiones cada vez más subdensas debido a una gravedad efectiva cada vez más expansiva. Las subdensidades se expanden mientras la región interna se vacía, formando en algunos casos una cáscara sobredensa que define los límites del vacío. De acuerdo a este simple escenario, podemos decir que la formación de los vacíos, en esencia, se debe a su naturaleza primordial subdensa y que, en contraste con el colapso de las sobredensidades de materia, es un proceso que se da desde adentro hacia afuera.

Las subdensidades, lejos de estar aisladas, se encuentran inmersas en una distribución compleja de materia, donde otras inhomogeneidades van creciendo por inestabilidad gravitatoria. A medida la expansión continúa y los vacíos aumentan de tamaño, estos eventualmente se enfrentan a sus pares en expansión. Por último, los vacíos estarán limitados por un déficit mínimo de densidad. Mientras las sobredensidades pueden crecer, en principio, indefinidamente, las subdensidades tienen un límite definido por $\delta_{min} = -1$ (ver Ec. 1.38). De este punto de vista, la influencia de las inhomogeneidades externas tienen un papel dominante en la evolución de los vacíos.

2.3. Alineaciones galácticas

La elección del alineamiento como la propiedad a estudiar de galaxias en vacíos para esta Tesis surge de forma natural. Los vacíos, por las características de su definición e identificación, poseen una dirección privilegiada: la dirección radial. Es directo pensar en el estudio de la alineación de las galaxias respecto de esta dirección privilegiada ya que se relaciona con el ambiente del vacío cósmico en sí, que es un componente dominante de la Red Cósmica, como también con propiedades dinámicas y astrofísicas de las galaxias. Como consecuencia del conjunto de procesos físicos que culminan en la formación y evolución de las galaxias, sus formas y orientaciones y las estructuras en gran escala en las que están inmersas pueden tener una coherencia significativa, y las propiedades de estas galaxias pueden diferir de las esperadas para galaxias orientadas al azar (e.g., véase [Mo et al. 2006](#)). Así, los estudios de las señales de alineación intrínseca permiten explorar los vínculos entre la evolución conjunta de las galaxias y sus estructuras circundantes (e.g. [Panko et al., 2013](#), y sus referencias). Teniendo en cuenta estos hechos, el análisis de las orientaciones de las galaxias en el contexto tanto del entorno local como de las estructuras en gran escala puede ser crucial para probar los escenarios de formación y evolución de las galaxias, en particular para predicciones teóricas de su momento angular (e.g., [Peebles, 1969](#)).

Para una muestra suficientemente grande de galaxias en un universo homogéneo e isótropo cabría esperar que las propiedades galácticas, como las orientaciones y las elipticidades, fueran aleatorias. Por esta razón, cualquier orientación preferida neta detectada con respecto a una dirección determinada, cualquier correlación no nula entre las alineaciones de las galaxias, o cualquier otro fenómeno que indique una violación local de la isotropía, suele relacionarse con las fuerzas gravitatorias de marea que actúan sobre las galaxias en diferentes etapas evolutivas ([Doroshkevich, 1970](#), [Peebles, 1969](#), [White, 1984](#)). Además, los modelos de efectos de lentes gravitacionales, que explican las distorsiones aparentes coherentes en las imágenes de las galaxias y ayudan a restringir los modelos cosmológicos, están a su vez restringidos por lo bien que entendemos otras posibles fuentes de alineación coherente subyacente (e.g. [Codis et al., 2015](#), [Croft y Metzler, 2000](#), [Heavens et al., 2000](#), [Hirata y Seljak, 2004](#)).

En el caso de las galaxias espirales (ver [Fig. 1.1](#)), la distribución espacial de las estrellas en un disco define un plano preferente cuya normal está orientada aproximadamente hacia el eje de rotación. El campo de marea ejercido por re-

giones caracterizadas por estructuras como cúmulos, filamentos o vacíos, están presentes durante una parte considerable de la evolución de las galaxias, y podría producir rasgos observables en su vector de momento angular, o espín, original. El hecho de que las galaxias giren es indicativo de las condiciones físicas en las que se formaron, y la rotación en sí misma es ciertamente una prueba importante de cualquier teoría sobre el origen de las galaxias (Peebles, 1969). En general, se cree que el momento angular de las galaxias surge de la torsión gravitacional debida a la desviación del tensor de cizalla (o tensor de *shear*, en inglés) gravitacional y el tensor de inercia en las primeras etapas de formación (Doroshkevich, 1970, White, 1984). Así, el campo de espines galácticos contiene información sobre el campo de cizalla gravitacional y puede utilizarse, por ejemplo, para una reconstrucción estadística del mismo (Lee y Pen, 2000).

En esta Tesis definimos y probamos estadísticos robustos para el estudio de alineamientos (Cap. 5, basado en el trabajo Dávila-Kurbán et al. 2022) y estudiamos la orientación de los espines galácticos dentro y en los límites (llamados envolturas o cáscaras) de los vacíos con respecto a su dirección radial, en la simulación TNG300-1 (Cap. 6, basado en el trabajo Dávila-Kurbán et al. 2022). A continuación repasamos métodos y resultados importantes en este estudio con el fin de delinear la diversidad de resultados y el posible rol que juegan en éstos los métodos empleados.

2.3.1. Revisión de métodos y resultados

Las señales de alineación son algo esquivas, dada la variedad de direcciones preferidas que surgen de la distribución real de las estructuras circundantes y el hecho de que las orientaciones de las galaxias con respecto a cualquier dirección son principalmente aleatorias. Además, estas pueden ser alteradas debido a interacciones (ver, por ejemplo, Gómez et al., 2017a,b). Por estas razones, un método estadístico robusto para detectar y evaluar las señales de alineación y su importancia es una herramienta clave en los estudios de alineaciones entre galaxias y la distribución en gran escala de estructuras.

Alineamientos en catálogos observacionales

En observaciones, la búsqueda de la alineación galáctica se ha analizado en el contexto de estructuras que, en una medida razonable, pueden describirse con

simetría esférica, como los cúmulos de galaxias o los vacíos. El aspecto observacional de este tema de estudio tiene sus propias dificultades, principalmente el pequeño tamaño de las muestras y los efectos de proyección en la línea de visual.

Los primeros trabajos se centraron en la orientación de las galaxias con respecto al Supercúmulo Local y a otros cúmulos como Virgo y Coma (por ejemplo Godłowski 1993, Godłowski y Ostrowski 1999, Godłowski 1994, Hu et al. 1995, 1998, Kashikawa y Okamura 1992, Wu 2006, Yuan et al. 1997). Muchos de estos trabajos se basaban en el método “ángulo de posición (AP) – inclinación” (Flin y Godłowski, 1986, Jaaniste y Saar, 1978). En este método, los APs medidos de las galaxias (normalmente en placas fotográficas) se convierten en vectores tridimensionales utilizando ángulos de inclinación obtenidos a partir de las relaciones axiales proyectadas, b/a , donde b es el semieje menor proyectado y a el mayor. La distribución de estos vectores puede entonces compararse con una hipótesis nula, por ejemplo una distribución espacial isotrópica, y así evaluar si los datos son isotrópicos o anisotrópicos por comparación. Sin embargo, la forma de estas distribuciones isotrópicas puede verse afectada significativamente por los criterios de selección de las muestras (Aryal et al., 2000). Estos efectos pueden ser grandes cuando la muestra se selecciona a partir de conjuntos de datos incompletos (por ejemplo, una porción limitada del cielo) y conducen a estructuras artificiales en los datos. Por lo tanto, un método estadísticamente robusto que describa de forma fiable no sólo los datos, sino también la muestra de comparación, es de crucial importancia en estos análisis para concluir a favor de la isotropía o anisotropía en los datos.

Otros estudios observacionales emplean métodos similares que también se basan en estadísticas de discretización como el recuento normalizado de pares, $P(\cos(\theta))$, en intervalos del ángulo θ medido entre el objeto de interés y una dirección privilegiada determinada por alguna estructura como los centros de los cúmulos o los elementos de la Red Cósmica (por ejemplo Brainerd, 2005, Yang et al., 2006, Zhang et al., 2015). La importancia de estas estadísticas suele evaluarse mediante la comparación con un gran número de simulaciones de Monte Carlo con hipótesis de no alineamiento.

Varela et al. (2012, en adelante V12) realizaron una evaluación rigurosa de un modelo analítico para la distribución de θ , y su comportamiento en el caso isotrópico para la estimación de la significación estadística, basándose en trabajos anteriores (Betancort-Rijo et al., 2009, Brunino et al., 2007, Cuesta et al., 2008,

Lee, 2004). Este trabajo abordó algunas discrepancias que surgieron en estudios observacionales previos sobre la alineación de galaxias alrededor de vacíos, a saber Trujillo et al. (2006) y Slosar y White (2009) (en adelante T06 y S09, respectivamente). Lo que estos trabajos tienen en común es el uso del mismo buscador de vacíos por parte de Patiri et al. (2006), que busca las mayores esferas no solapadas dentro del volumen del relevamiento desprovistas de galaxias por encima de un determinado umbral de brillo. Sin embargo, hubieron diferencias significativas en la selección de las muestras de galaxias y en los métodos de medición de sus espines. Por ejemplo, T06 se limitó a seleccionar sólo galaxias con su disco de borde y de cara, mientras que V12 ajustó un modelo de disco grueso a todas las galaxias que fueron clasificadas como espirales por GalaxyZoo (Lintott et al. 2008). En particular, T06 analizó 201 galaxias de cara y de borde utilizando datos del SDSS-DR3 y del 2dFRGS (Colless et al., 2001) y encontró una tendencia significativa del espín de galaxias a estar en la dirección perpendicular a la dirección radial del vacío. Por otro lado, S09, utilizando dos muestras de 578 y 258 galaxias del SDSS-DR6 con criterios de selección similares, no encontró evidencia estadística a favor de alguna orientación preferencial. V12, utilizó el SDSS-DR7 y un procedimiento estadístico lo suficientemente robusto como para superar el problema de la indeterminación de la inclinación real de las galaxias calculada a partir de su relación axial aparente, y evaluaron la validez del procedimiento con extensas simulaciones de Monte Carlo. Los autores detectan una tendencia estadísticamente significativa de que las galaxias alrededor de los grandes vacíos (con radios de más de $15 h^{-1}\text{Mpc}$) tienen sus momentos angulares alineados con la dirección radial de los mismos.

El escenario estándar de la teoría del torque tidal (Lee et al. 2007, Lee y Pen 2000, 2001) postula una alineación neta preferente del vector de espín con el eje principal intermedio del tensor de *shear*. Este eje, a su vez, se encontraría ubicado tangencialmente a la superficie del vacío¹. De acuerdo con esta predicción, T06 encontró un nivel de confianza del 99.7% en que las galaxias espirales situadas en las cáscaras de vacíos cósmicos de radio mayor a $10 \text{ Mpc } h^{-1}$ tienen ejes de rotación que se encuentran preferentemente en la superficie del vacío. SW09 no encontró ninguna evidencia estadística de desviación de orientaciones aleatorias; argumentan que los resultados de T06 podrían ser posiblemente una fluctuación

¹Esta comparación es razonable pero no válida siempre ya que el cálculo del tensor de *shear* y la identificación de paredes y/o vacíos pueden llevarse a cabo con distintos métodos que pueden no corresponderse directamente entre sí.

estadística dado que el catálogo utilizado en SW09 es considerablemente mayor y aumenta drásticamente el número de vacíos. Por otro lado, V12 consideró vacíos con radios mínimos de 15 Mpc h^{-1} y encontró una señal significativa ($> 98.8\%$) de que la alineación del espín de las galaxias alrededor de estos vacíos sea preferentemente paralela al radio-vector, mientras que para vacíos más pequeños esta tendencia desaparece y los resultados son consistentes con ninguna alineación especial. Además, V12 también encuentra que la fuerza de la alineación depende de la distancia entre las galaxias y la superficie del vacío y que, independientemente del tamaño del vacío, para galaxias más lejanas que aproximadamente 5 Mpc h^{-1} no hay una dirección preferente en la distribución de las alineaciones. En cuanto al desacuerdo con la orientación perpendicular neta (T06) o sin orientación (SW09), V12 argumenta que el pequeño tamaño de la muestra de galaxias alrededor de los vacíos con $R \geq 10 \text{ Mpc h}^{-1}$ utilizada en estos trabajos podría enmascarar la señal de alineación que ellos encuentran.

El escenario de las alineaciones de espín con las láminas (o paredes) sigue sin estar claro. Utilizando observaciones basadas en datos de placas fotográficas, Lee y Pen (2002) y Lee et al. (2007) concluyeron que los espines de las galaxias tienden a situarse dentro de las láminas, mientras que utilizando datos del SDSS, Tempel y Libeskind (2013) y Zhang et al. (2015) encontraron que los momentos angulares de las galaxias apuntan preferentemente perpendiculares al plano de la lámina, aunque con una señal débil en ambos tipos de alineación; Tempel y Libeskind (2013) encuentra esta leve señal en galaxias elípticas mientras que Zhang et al. (2015) lo hace para galaxias en entornos laminares. Estos últimos resultados parecen coherentes con el resultado de V12 para vacíos, sin embargo, los resultados basados en simulaciones coinciden en general en que los momentos angulares (de halos y, con menor señal, galaxias) apuntan preferentemente paralelos a las estructuras planas (e.g, Libeskind et al. 2013), y esta tendencia parece hacerse más fuerte con los halos más masivos (e.g, Forero-Romero et al. 2014).

Alineamientos de halos y galaxias en simulaciones

Se suele suponer que, durante las primeras etapas de formación, la materia bariónica y la materia oscura compartieron una evolución similar y probablemente adquirieron el mismo momento angular específico antes de la formación del disco galáctico (e.g., Fall y Efstathiou, 1980). El estudio de los halos de materia oscura alineados fue posible, y se convirtió en un tema popular, después de que las

simulaciones de N -cuerpos tuvieran suficiente resolución para realizar estudios de esta naturaleza (e.g., Cuesta et al., 2008, Forero-Romero et al., 2014, Joachimi et al., 2015, Kiessling et al., 2015, Libeskind et al., 2013, y sus referencias).

Los métodos de detección de la alineación y los resultados obtenidos son diversos. Por ejemplo, Forero-Romero et al. (2014) estudiaron la alineación de la forma, el momento angular y la velocidad peculiar de los halos de materia oscura con respecto a la Red Cósmica, tal como se describe utilizando el campo de marea o la cizalla de velocidad, empleando la simulación Bolshoi (Riebe et al., 2013). Los autores cuantifican los alineamientos midiendo la fracción de halos que se alinea preferentemente con uno de los autovectores en la definición local de la Red Cósmica, y con el valor medio del ángulo entre un autovector y el vector de interés. Encontraron la mayor alineación en las formas de los halos con respecto a filamentos y paredes definidas por el campo de marea, pero cuando se definen por la cizalla de velocidad encontraron una anti-alineación (alineación perpendicular) de estos con los halos masivos. Para el momento angular, sólo encontraron una señal más débil para los halos más masivos de estar anti-alineados con filamentos, y de estar alineados a lo largo de las estructuras laminares de la cizalla de velocidad. Estos resultados, sin embargo, discrepan con trabajos anteriores (Aragon-Calvo, 2013, Aragón-Calvo et al., 2007, Hahn et al., 2007) que sí detectan alineaciones para los halos menos masivos. Forero-Romero et al. (2014) argumentan que esto podría deberse a que la señal de alineación es altamente sensible a la descripción de la Red Cósmica en pequeña escala. Además, encuentran que las velocidades peculiares son preferentemente paralelas a las paredes y a los filamentos. Sus resultados indican que las propiedades de alineación de los halos de materia oscura pueden depender de la definición física de la Red Cósmica: las definiciones a partir del campo de marea y de la cizalla de velocidad aportan información complementaria.

En los últimos años, a medida que se dispone de simulaciones con una resolución cada vez mayor, se han realizado varios estudios sobre las alineaciones de los espines, así como sobre las formas de las galaxias o halos con respecto a las distintas subestructuras de la Red Cósmica. Aunque podría ser tentador pensar que las cáscaras de los vacíos cósmicos, en las que se centra este trabajo, son equivalentes a las subestructuras de láminas de la Red Cósmica, hay que tener en cuenta que los algoritmos prácticos para identificarlas son significativamente diferentes (véase, por ejemplo, la revisión Joachimi et al. 2015). Codis et al. (2018) y

Kraljic et al. (2019), por ejemplo, estudian la distribución de los ángulos medidos entre el espín de galaxias y halos y los diferentes elementos circundantes de la Red Cósmica en las simulaciones Horizon–AGN (Dubois et al., 2014) y SIMBA (Dave et al., 2019), respectivamente, utilizando el código público DISPERSE² (Sousbie 2013) para identificar subestructuras en las simulaciones. Sus resultados coinciden en la detección de un “spin-flip” dependiente de la masa de galaxias y halos: es más probable que el espín de las galaxias de baja masa se sitúe en el plano de las láminas, mientras que las galaxias masivas tienen un espín preferentemente perpendicular a las mismas.

Comentarios finales

Podemos pensar en varias razones para esta diversidad resultados no concluyentes. Desde el punto de vista observacional, una de las principales dificultades sigue siendo la de relacionar las formas observadas con las direcciones de espín o las orientaciones elipsoidales, utilizando métodos diferentes y dando posiblemente resultados distintos con la misma muestra. La propia señal, además, parece depender significativamente de los parámetros utilizados para seleccionar las submuestras de galaxias (como la luminosidad, la morfología, los colores, etc.), lo que obstaculiza una comparación clara entre los distintos trabajos. Además, el pequeño número de galaxias en los vacíos implica que los resultados de las observaciones en estos entornos tienen grandes incertidumbres estadísticas (véase Zhang et al. 2015). Por estas razones, además de la mejora de la potencia de cálculo, los estudios basados en la simulación han prosperado en este ámbito, especialmente en la búsqueda de alineaciones de los halos bariónicos y de materia oscura con respecto a los componentes de la estructura en gran escala (véase, por ejemplo, Codis et al. 2018 y Kraljic et al. 2019).

Las diferencias en los resultados pone de manifiesto la importancia, no sólo de un mayor tamaño de la muestra, sino también del uso de métodos estadísticos robustos y fiables para evaluar con alta confianza estadística la validez de la señal de alineación detectada.

²<http://www.iap.fr/users/sousbie/disperse.html>

Capítulo 3

Estadísticos para el estudio de la distribución de materia

El análisis de la distribución espacial en gran escala de las galaxias ha sido decisivo para conformar nuestra comprensión actual de la evolución del Universo (por ejemplo Alam et al., 2017, Cole et al., 2005, Davis y Peebles, 1983, eBOSS Collaboration et al., 2021, Efstathiou et al., 2002, Eisenstein et al., 2005, Feldman et al., 1994, Tegmark et al., 2004). Dada la naturaleza estocástica de esta distribución, estos análisis requieren herramientas estadísticas robustas para extraer eficazmente la información cosmológica codificada en los estudios de galaxias. Las herramientas más utilizadas para caracterizar la estructura en gran escala del Universo son las estadísticas de dos puntos, como el espectro de potencia, $P(\mathbf{k})$ (Ec. 1.39), y su transformada de Fourier, la función de correlación bipuntual $\xi(\mathbf{r})$. Sin embargo se pueden estudiar estadísticos de órdenes superiores, es decir, estadísticos de tres, cuatro, o más puntos. Éstos son notoriamente complicados de calcular pero contienen información que podría ser crucial para completar nuestro entendimiento de la distribución de galaxias en gran escala (Croton et al., 2004a,b).

En este capítulo vamos a definir brevemente conceptos básicos del estadístico más tradicional, la función de correlación bipuntual, en la Sec. 3.1. Recapitulamos información disponible en la literatura acerca de estadísticos relacionados a los órdenes superiores de agrupación en la Sec. 3.3. Finalmente, en la Sec. 4, basada en el trabajo de Dávila-Kurbán et al. (2021), proponemos un método para disminuir la varianza y reducir el costo computacional en el cálculo de la función de correlación bipuntual.

3.1. Función de correlación bipuntual (2PCF)

La medida cuantitativa más utilizada de la estructura en gran escala es la función de correlación bipuntual de las galaxias (2PCF), $\xi(r)$, que traza la amplitud de la agrupación de galaxias en función de la escala (r) (Davis y Peebles, 1983, Davis et al., 1988, Fisher et al., 1994, Loveday et al., 1994, Norberg et al., 2001, Zehavi et al., 2001). Definimos primero la 2PCF junto con algunos de sus estimadores más populares siguiendo el enfoque de Coil (2013), para luego analizar los efectos del *redshift* en la correlación en la sección siguiente.

La función de correlación bipuntual es una medida del exceso de probabilidad, dP , respecto del esperado para una distribución aleatoria de Poisson (i.e. sin ninguna agrupación intrínseca), de encontrar una galaxia en un elemento de volumen dV separado de otra galaxia por una cantidad r ,

$$dP = n[1 + \xi(r)]dV, \quad (3.1)$$

donde n es la densidad numérica media de la muestra de galaxias en cuestión (Coil, 2013, Peebles, 1980). Valores más grandes de $\xi(r)$ implican un agrupación de mayor amplitud de las galaxias en la escala r , mientras que valores negativos de $\xi(r)$ indicarían una anti-correlación. Generalmente, las mediciones de $\xi(r)$ se realizan en el espacio comóvil, donde r tiene unidades de $h^{-1}\text{Mpc}$. Para medir $\xi(r)$, se cuentan pares de galaxias como función de la separación entre ellas y se divide por la cantidad esperada para una distribución sin agrupamiento intrínseco o distribución aleatoria. Para ello hay que construir un “catálogo aleatorio” de comparación constituido de puntos con posiciones independientes entre sí, de manera tal que tenga una cobertura tridimensional idéntica a la de los datos. De esta manera, se utiliza el cociente entre pares de galaxias observadas en los datos y pares de puntos del catálogo aleatorio para estimar $\xi(r)$. Se han propuesto y probado diversos estimadores para $\xi(r)$, siendo uno de los primeros estimadores en ser utilizados ampliamente el de Davis y Peebles (1983):

$$\xi_{DP}(r) = \frac{n_R}{n_D} \frac{DD(r)}{DR(r)} - 1, \quad (3.2)$$

donde DD es el recuento de pares de galaxias en cada intervalo de separación r en el catálogo de datos y DR el conteo análogo pero entre los catálogos de datos y el aleatorio. Este cociente está normalizado por las cantidades n_D y n_R que son

las densidades numéricas medias de galaxias en los catálogos de datos y aleatorio respectivamente. Luego, Hamilton (1993a), utilizando una cantidad RR definida como el conteo de pares de galaxias en función de la separación r en el catálogo aleatorio, introdujo un estimador con menores errores estadísticos,

$$\xi_H(r) = \frac{DD(r) RR(r)}{(DR(r))^2} - 1. \quad (3.3)$$

El estimador más popularmente utilizado fue presentado por Landy y Szalay (1993):

$$\xi_{LS}(r) = \frac{1}{RR(r)} \left[DD(r) \left(\frac{n_R}{n_D} \right)^2 - 2DR(r) \left(\frac{n_R}{n_D} \right) + RR(r) \right] - 1. \quad (3.4)$$

Este estimador requiere más tiempo de cálculo que el estimador de Hamilton (Ec. 3.3). Sin embargo es menos sensible al tamaño del catálogo aleatorio y disminuye errores de borde que pueden afectar a las mediciones de agrupamiento en gran escala (Coil, 2013, Kerscher et al., 2000).

Puede verse de la forma de los estimadores indicados aquí que la estimación de $\xi(r)$ es significativamente sensible a qué tan fielmente refleja el catálogo aleatorio los efectos de selección espacial y de *redshift* en los datos. El catálogo aleatorio también debe ser lo suficientemente grande como para no introducir errores poissonianos en el estimador. En la Sec. 4 planteamos un tratamiento de estos catálogos aleatorios para reducir su ruido poissoniano y así reducir la varianza en el cálculo de funciones de correlación bipuntuales.

3.2. Función de correlación anisotrópica

3.2.1. 2PCF en el espacio de *redshift*

En esta subsección, siguiendo el enfoque de Coil (2013), resumiremos los efectos que tiene el *redshift* en las mediciones de correlación introducidas en la sección anterior.

Para calcular la función de correlación bipuntual en catálogos observacionales, el dato que tenemos para estimar la ubicación de las mismas en la línea de la visual no es su distancia, sino su *redshift*. Esto implica la necesidad de elegir un modelo cosmológico que nos permita relacionar el *redshift* medido con la distancia

estimada. No obstante, los errores introducidos por la suposición de una cosmología dada no son generalmente dominantes, de modo que, aunque en teoría se podrían asumir diferentes parámetros cosmológicos y comprobar qué resultados son consistentes con los valores asumidos, esta labor no es generalmente necesaria. Por otro lado, debe tomarse en cuenta que las velocidades peculiares de las galaxias introducen distorsiones en la línea de la visual en el espacio de *redshift*. Las mediciones de estas distorsiones pueden restringir parámetros cosmológicos y también pueden integrarse para recuperar la función de correlación del espacio real subyacente (Coil, 2013, Sargent y Turner, 1977).

En escalas espaciales pequeñas, es decir, escalas típicas de sobredensidades virializadas colapsadas tales como grupos y cúmulos, las galaxias tienen grandes movimientos aleatorios entre sí. Esto significa que aunque todas las galaxias del grupo o cúmulo tienen una distancia física similar al observador, pueden presentar un amplio rango de valores de *redshift*. El resultado de esto es un alargamiento a lo largo de la visual en los mapas del espacio de *redshift* dentro de estas regiones sobredensas. Este efecto se conoce como “dedos de dios” (FOG por sus siglas en inglés, *fingers of god*). El efecto resultante de los FOG es que los grupos y cúmulos parecen extenderse radialmente hacia el observador, i.e. a lo largo de la línea de visual.

Este efecto de extensión radial puede verse en el panel inferior izquierdo de la Fig. 3.1 (Tegmark et al., 2004), que muestra galaxias en el espacio de *redshift* “extendiéndose” hacia el observador, mientras que en el panel inferior derecho estos efectos han sido modelados y eliminados. Las escalas mayores sufren otro tipo de distorsiones en el espacio de *redshift* debido a que las galaxias que están comenzando una interacción gravitatoria con estructuras en colapso presentan movimientos de flujo coherente (Kaiser, 1987). Estas galaxias adyacentes se mueven todas en la misma dirección, hacia la región de colapso, lo que provoca una distorsión de contracción (al contrario de los FOG) en la línea de la visual del espacio de *redshift*.

Estadísticos bidimensionales de la agrupación de galaxias pueden mostrar claramente las distorsiones del espacio de *redshift*. A pesar de la utilidad de estas distorsiones, ya que contienen información sobre la densidad de materia subyacente y los movimientos de las galaxias, éstas impiden medir de forma directa la correlación bipuntual en el espacio real. Tenemos entonces que, en lugar de $\xi(r)$ se mide $\xi(s)$ en el espacio de *redshift*, donde $s = |\mathbf{s}|$ es la separación entre un

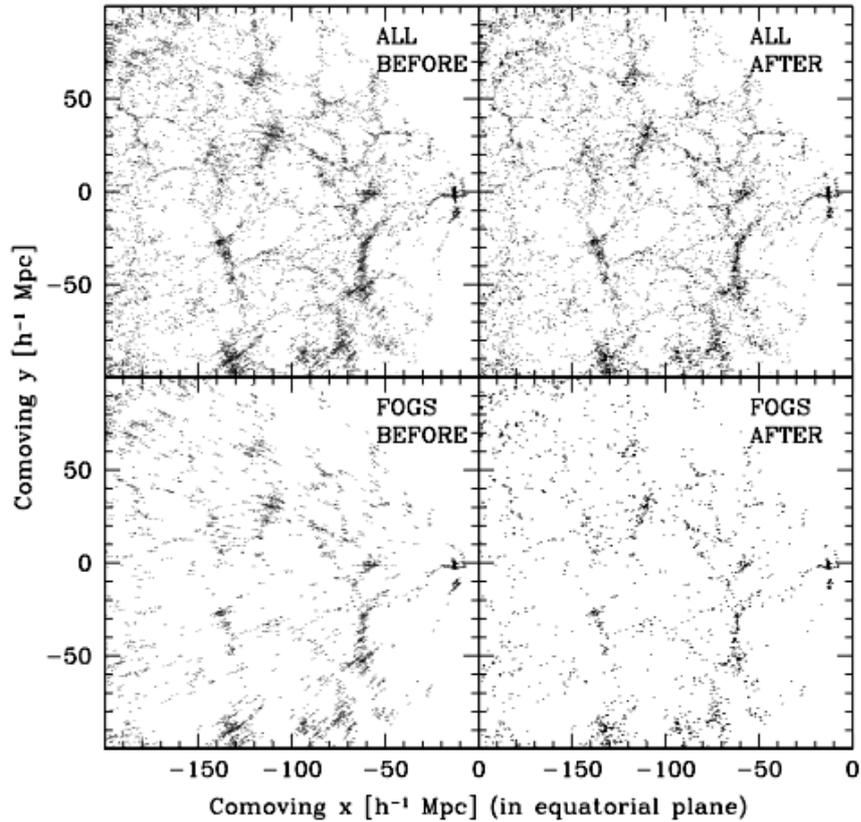


Figura 3.1: Ilustración de los “dedos de dios” (FOG), donde puede apreciarse la elongación de las estructuras virializadas a lo largo de la línea de visual. Se muestran galaxias de una rodaja del SDSS en un espacio bidimensional comóvil. Los paneles superiores muestran todas las galaxias de esta rodaja mientras que los inferiores muestran sólo galaxias que han sido identificadas como afectadas por los FOG. Los paneles de la derecha muestran la posición de estas galaxias luego de modelar y eliminar los efectos de los FOG, con el observador en el cero de ambos ejes. Los efectos de los FOG pueden observarse en el panel inferior izquierdo donde las galaxias se dispersan en la dirección de la visual. Figura original de Tegmark et al. (2004), extraída de Coil (2013).

par de galaxias en dicho espacio. La comparación entre los resultados del cálculo de $\xi(s)$ para diferentes muestras de galaxias y diferentes *redshifts* no es directa, ya que la amplitud de las distorsiones en este espacio difieren según el tipo de galaxia y su *redshift* medido. Además, las distorsiones en pequeñas y grandes escalas disminuyen la amplitud de la agrupación en comparación con las escalas intermedias, por lo que la $\xi(s)$ no sigue una ley de potencia en las mismas escalas que $\xi(r)$ (Coil, 2013).

Si queremos conocer la agrupación física de las galaxias, independiente de las velocidades peculiares de estas, debemos calcular $\xi(r)$. Ahora bien, para recuperar la función de correlación en el espacio real, se puede medir ξ de forma bidimensional: tanto perpendicularmente como a lo largo de la línea de visual. Las separaciones entre las dos galaxias, tanto a lo largo (r_π) como perpendicular (r_σ) a la línea de visual pueden definirse como: $r_\pi = (\mathbf{s} \cdot \mathbf{l})/|\mathbf{l}|$ y $r_\sigma = \sqrt{\mathbf{s} \cdot \mathbf{s} - \pi^2}$, donde \mathbf{s} es la separación entre galaxias en el espacio de *redshift* y \mathbf{l} es la distancia media al par. Otras nomenclaturas pueden ser encontradas en la literatura donde se denominan a las distancias perpendicular y paralela, respectivamente, como $[\sigma, \pi]$ o bien $[r_\sigma, r_\pi]$, entre otras. También es común expresar la 2PCF bidimensional, como lo haremos en secciones siguientes, en términos de la separación $s = |\mathbf{s}|$ y el coseno del ángulo entre la línea de la visual y \mathbf{s} , denotado con μ : $\xi(s, \mu)$.

En la Fig. 3.2 (Reid et al., 2014) se muestra un ejemplo de una medición de $\xi(r_\sigma, r_\pi)$. En un universo homogéneo, isotrópico y sin distorsiones, las características visibles en esta figura serían circulares. En escalas pequeñas los contornos se distorsionan y muestran los “dedos de dios” a lo largo de la línea de la visual, correspondientes a las galaxias en sobredensidades virializadas. En escalas menores a $8h^{-1}\text{Mpc}$ aproximadamente, los contornos se extienden en la dirección r_π mientras que en escalas más grandes se aplanan debido al “efecto Kaiser” (Kaiser, 1987), i.e. para esta muestra de galaxias, esta es la escala aproximada en la que el cuadrupolo transiciona entre valores negativos y positivos (ver subsección siguiente). Esto es indicativo del movimiento de flujo coherente de las galaxias hacia estructuras en colapso gravitatorio.

La amplitud de este efecto, al ser este causado por la caída de galaxias sobre estructuras masivas en formación, depende de Ω_m (Ec. 1.14). La anisotropía a gran escala en el plano de la función de correlación bidimensional depende de $\beta \equiv \Omega_m/b$ en escalas lineales (Kaiser, 1987), donde b es el sesgo o *bias* definido como el cociente entre las fluctuaciones de densidad en la población de galaxias

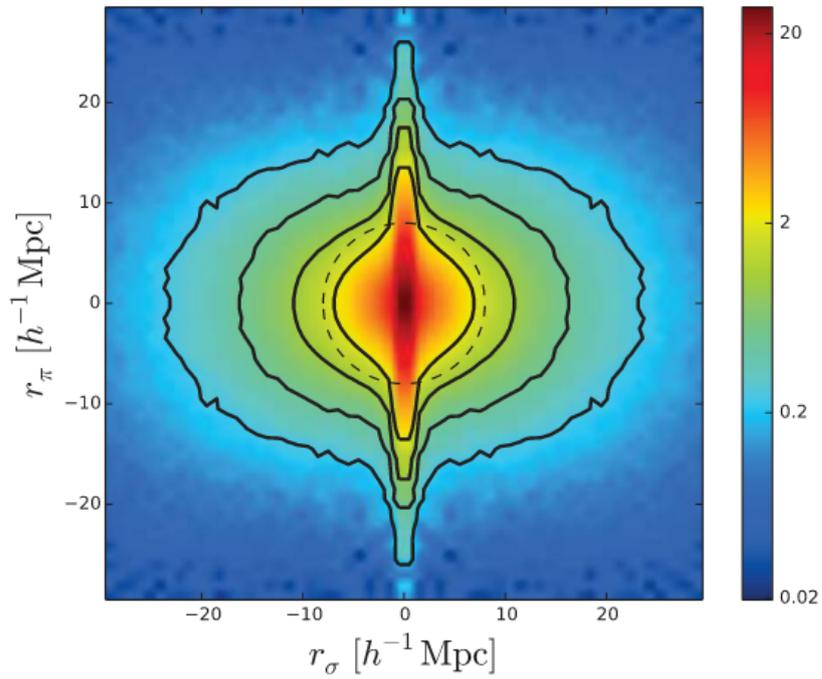


Figura 3.2: La función de correlación bidimensional del espacio de *redshift* de galaxias del SDSS-III CMASS (siglas de *Constant (stellar) MASS* en inglés). Se muestran las distorsiones a lo largo de la línea de la visual (eje Y) en el espacio de *redshift* causadas por velocidades peculiares, con isocontornos negros para valores $\xi = [2, 1, 0.5, 0.25]$. Se muestra en línea negra de trazos la escala aproximada para esta muestra ($\simeq 8h^{-1}\text{Mpc}$) de la transición entre los efectos FOG y Kaiser. Figura extraída de Reid et al. (2014)

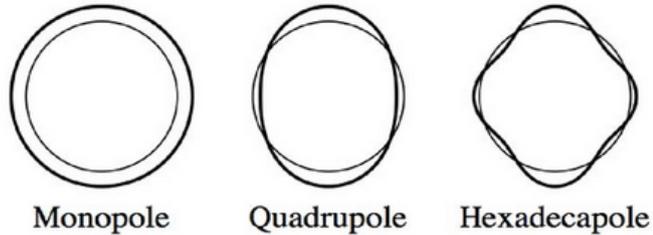


Figura 3.3: Formas de los primeros tres armónicos pares de Legendre ($\ell = 0, 2, 4$), de Hamilton (1998), dibujados como distorsiones en una simetría circular. El monopolo cuantifica la amplitud de la distorsión mientras que el cuadrupolo y el hexadecapolo cuantifican la forma de la distorsión.

y en la materia oscura. Esta utilidad para restringir parámetros cosmológicos es la razón por la que las distorsiones del espacio de *redshift* son de gran interés en la cosmología, mientras que aporta información de índole más dinámica para estudios centrados en la estructura en gran escala.

3.2.2. Multipolos de Legendre

Las anisotropías de la función de correlación se suelen cuantificar mediante los multipolos de $\xi(r_\sigma, r_\pi)$, definidos como

$$\xi_\ell(s) = \frac{2\ell + 1}{2} \int \xi(r_\sigma, r_\pi) L_\ell(\cos(\theta)) d\cos(\theta), \quad (3.5)$$

donde L_ℓ son los polinomios de Legendre, y θ es el ángulo entre s y la línea de visual. El monopolo, cuadrupolo y hexadecapolo, i.e. $\ell = 0, 2, 4$, cuantifican las distorsiones en el espacio de *redshift* como se muestra de forma esquemática en la Fig. 3.3, donde generalmente se encuentra que $\xi_{\ell=4} \simeq 0$ en un amplio rango de escalas (e.g. Gil-Marín et al., 2020). En la teoría lineal el parámetro β mencionado en la subsección anterior puede extraerse del cociente entre los momentos cuadrupolares y monopulares de la función de correlación anisotrópica, ξ_2/ξ_0 , (Hamilton, 1998). La descomposición de la función de correlación anisotrópica en términos de estos armónicos resulta muy conveniente, dado que podemos obtener toda la información contenida en la función completa comprimida en estas tres funciones unidimensionales: $\xi_{\ell=0,2,4}$.

3.3. Estadísticos de agrupamiento de alto orden

Las estructuras que forman las galaxias en gran escala contienen vasta información acerca de la evolución tanto lineal como no lineal del agrupamiento de galaxias. La naturaleza de dicho agrupamiento depende de numerosos efectos tanto en escalas grandes como pequeñas, tales como los parámetros cosmológicos, efectos e historia ambientales de las galaxias y cúmulos, y la forma en la que las componentes luminosas y oscuras del Universo se acoplan y evolucionan. Estudiando los distintos órdenes del agrupamiento de galaxias se espera arrojar luz a los procesos físicos que afectan el agrupamiento (Croton et al., 2004b).

La función de correlación bipuntual, definida en la Sec. 3.1, ha sido tradicionalmente la herramienta más usada para analizar dichas distribuciones, proveyendo una descripción del agrupamiento en órdenes bajos. Sin embargo, a pesar de su utilidad, la función de correlación bipuntual sólo provee una descripción completa del agrupamiento en el caso de una distribución gaussiana. Una explicación más completa del agrupamiento debe incluir funciones de correlación de más altos ordenes, aunque estas pueden ser difíciles de extraer (ver Croton et al. 2004a y Baugh et al. 2004 para un análisis del agrupamiento galáctico hasta sexto orden).

En este contexto, investigadores han explorado otras estadísticas de agrupamiento para obtener información de alto orden de una distribución de galaxias. Las medidas de alto orden incluyen la relación de densidad *genus-threshold* (Gott et al., 1989, Ryden et al., 1989), momentos de recuentos en celdas (e.g. Saunders et al., 1991), estadísticas de percolación (Shandarin y Zeldovich, 1983), y la distribución de probabilidad de los recuentos en celdas (CIC por sus siglas en ingles, *counts-in-cells*) (Vogeley et al., 1994). Aquí consideramos las estadísticas de vacíos en la distribución de galaxias, que son un caso especial del análisis de recuentos en celdas.

En el marco de la distribución de estadísticos CIC, el nombre “vacío” indica la ausencia absoluta de galaxias en una dada celda, distinto al concepto de vacío cósmico -subdensidades en la estructura a gran escala- que hemos introducido en los primeros capítulos. En las secciones siguientes (Sec. 3.3-3.3.5) emplearemos el término “vacío” con la primera denominación, refiriéndonos a la ausencia de galaxias, a menos que se indique lo contrario.

Históricamente, muchos astrónomos han mostrado interés en usar estadísticas de vacío (e.g. Balian et al., 1989, Bouchet et al., 1993, Fry, 1986, Fry et al., 1989, Gaztanaga y Yokoyama, 1993, Maurogordato y Lachieze-Rey, 1987, Vogeley et al.,

1994). Este enfoque ha resultado provechoso en el sentido de que los resultados son obtenibles fácilmente y están bien fundamentados en un marco teórico sólido que relaciona directamente la distribución de vacíos a la distribución de galaxias de altos órdenes (Balian et al., 1989, Fry, 1986, White, 1979).

A continuación presentaremos conceptos fundamentales y necesarios para entender estadísticas de vacío (en particular la función de probabilidad de vacío, Sec. 3.3.1), haremos un repaso de resultados en la literatura obtenidos con ellas (Sec. 3.3.4), y finalizamos con una discusión de los mismos (Sec. 3.3.5).

3.3.1. La función de probabilidad de vacío (VPF)

En esta subsección y las siguientes seguimos el enfoque de Croton et al. (2004b) para presentar los conceptos básicos que se necesitan para definir la función de probabilidad de vacío, la noción de “escalamiento jerárquico” de los órdenes de correlación (Sec. 3.3.2), y finalmente un resumen de algunos modelos fenomenológicos populares de agrupamiento jerárquico (Sec. 3.3.3) que utilizaremos en capítulos posteriores de esta Tesis.

Para una dada distribución de galaxias, la función distribución de probabilidad de conteos (CPDF), $P_N(V)$, se define como la probabilidad de encontrar exactamente N galaxias en una celda de volumen V ubicada aleatoriamente dentro de la muestra. Para el caso donde $N=0$ tenemos una función de probabilidad de vacío (VPF), $P_0(V)$. Esta función depende de todas las funciones de correlación de p puntos (Fry, 1985, Otto et al., 1986, White, 1979) según la relación:

$$P_0(n, V) = \exp \left[\sum_{N=1}^{\infty} \frac{(-n)^N}{N!} \int_V w_N(\mathbf{x}_1, \dots, \mathbf{x}_N) d^3x_1 \dots d^3x_N \right], \quad (3.6)$$

donde n es la densidad numérica promedio de objetos, los w_N son las funciones de correlación de N puntos, y las x_i son las posiciones de los objetos dentro del volumen V . Por definición $w_1 = 1$, $w_2 = \xi_2$ (la función de correlación bipuntual), etc. En general, la Ec. 3.6 converge sólo para grandes valores de N (Fry, 1984, White, 1979), de modo que la VPF es sensible a agrupamientos de alto orden.

Eligiendo celdas esféricas para muestrear la distribución hacemos que P_0 sea una función dependiente sólo del radio R de la esfera, y entonces la relación entre la VPF las funciones de correlación de p puntos puede escribirse como ¹:

¹Notamos que, al necesitar la cantidad \bar{N} , optamos por usar el subíndice p en lugar de N por motivos de legibilidad.

$$P_0(R) = \exp\left\{\sum_{p=1}^{\infty} \frac{[-\bar{N}(R)]^p}{p!} \bar{\xi}_p(R)\right\}. \quad (3.7)$$

Aquí, \bar{N} es el número promedio de objetos en una celda de volumen V (nV), y $\bar{\xi}_p(R)$ es la función de correlación de orden p promediada sobre el volumen esférico V .

White (1979) también muestra que la VPF puede interpretarse como un generador de probabilidad de conteos en un volumen,

$$P_N(V) = \frac{(-n)^N}{N!} \frac{d^N}{dn^N} P_0(V). \quad (3.8)$$

Además, en el límite continuo, la VPF es la transformada de Laplace de la función de distribución de la densidad $f(n)$ (Fry, 1985),

$$P_0(V) = \int_0^{\infty} e^{-nV} f(n) dn. \quad (3.9)$$

Para una distribución completamente aleatoria (Poisson) se tiene $\xi_p \equiv 0$ para todo $p > 1$, de modo que P_0 se reduce a una expresión analítica más simple:

$$P_{0, \text{Poisson}}(R) = \exp[-\bar{N}(R)]. \quad (3.10)$$

Por lo tanto, si la VPF de una distribución de puntos muestra cualquier desviación de este comportamiento, se concluye a favor de la presencia de alguna clase de agrupamiento.

3.3.2. Escala jerárquica y la VPF reducida (RVPF)

El concepto de “escala jerárquica” puede generalizarse asumiendo que cada función de correlación de p puntos depende sólo del producto entre la función de correlación bipuntual y un coeficiente de escala, S_p , que es adimensional:

$$\bar{\xi}_p(R) = S_p \bar{\xi}^{p-1}(R), \quad (3.11)$$

donde hemos eliminado el subíndice 2 para la función de correlación bipuntual en el lado derecho por comodidad. Se denomina “estructura jerárquica” a tal organización de los ξ_p , y el modelo jerárquico determina los coeficientes de escala S_p . Tradicionalmente, $S_3 = \bar{\xi}_3/\bar{\xi}^2$ se denomina *asimetría* de la distribución y $S_4 = \bar{\xi}_4/\bar{\xi}^3$ como la *curtosis*.

La idea jerárquica es directamente aplicable a la VPF, ya que depende de una suma infinita de funciones de correlación de p puntos. La suposición jerárquica nos permite eliminar las funciones de correlación de orden superior de la Ec. 3.7:

$$P_0(R) = \exp \left[\sum_{p=1}^{\infty} \frac{[-\bar{N}]^p}{p!} S_p \bar{\xi}^{p-1}(R) \right]. \quad (3.12)$$

Además, la relación de escalamiento anterior nos permite expresar la VPF como una función de $\bar{N}\bar{\xi}$ solamente, donde la variable de escala $\bar{N}\bar{\xi}$ puede pensarse, a grandes rasgos, como el número medio de galaxias en una celda en exceso de lo esperado dada la densidad media de la muestra. Formalizamos esta idea considerando primero la expresión analítica de la VPF de una muestra puramente aleatoria (Eq. 3.10). Para la situación jerárquica, podemos definir un parámetro χ , con $P_0 = e^{\bar{N}\chi}$, denominado función de probabilidad de vacío reducida (RVPF por sus siglas en inglés, véase Fry, 1986):

$$\chi = -\ln(P_0)/\bar{N}. \quad (3.13)$$

Observamos aquí que, independientemente de la suposición jerárquica, χ normaliza la contribución de Poisson a la distribución, y es evidente que los efectos de la agrupación aparecerán como los valores $\chi < 1$. Combinando las Ecs. 3.12 y 3.13, la RVPF adopta la forma:

$$\chi(\bar{N}\bar{\xi}) = \sum_{p=1}^{\infty} \frac{S_p}{p!} (-\bar{N}\bar{\xi})^{p-1}. \quad (3.14)$$

Esto muestra la relación de escala mencionada, y la forma de $\chi(\bar{N}\bar{\xi})$ caracteriza así la distribución de los vacíos. Si se cumple la hipótesis de la relación de escala, esperamos que todas las muestras de galaxias de diferente densidad y amplitud de agrupación colapsen en una única curva, ya que todas son una función de la misma variable de escala. En otras palabras, si las RVPF de todas las muestras conforman una sola función, entonces los coeficientes S_p son independientes de la escala para $p > 2$, ya que al expresar la VPF como función de $\bar{N}\bar{\xi}$ estamos eliminando la dependencia con la amplitud de la 2PCF (Croton et al., 2004b, Fry, 1986, Vogelej et al., 1994). Por el contrario, la curva no será universal para diferentes rangos de magnitud de las galaxias si los coeficientes S_p dependen fuertemente de esta propiedad. Sin embargo, los valores de S_p dependen, en el mejor de los casos, sólo débilmente de la magnitud (Croton et al., 2004a).

La variable de escala puede definirse como $\bar{N}\bar{\xi} = 4\pi n J_3$ (ver Fry, 1986) y es aproximadamente el número de vecinos en exceso respecto al azar para cualquier galaxia, donde n es la densidad numérica media de las galaxias y J_3 es el segundo momento de la función de correlación bipuntual (Vogeley et al., 1994). En el escenario jerárquico, cuando $\bar{N}\bar{\xi} \ll 1$ siempre se recupera la VPF de Poisson, $\chi(\bar{N}\bar{\xi}) = 1$, independientemente del patrón de la agrupación real o su amplitud. En el régimen en el que $\bar{N}\bar{\xi} < 1$ vemos a partir de la Eq. 3.14 que la RVPF está dominada por la contribución gaussiana: $1 - \frac{1}{2}\bar{N}\bar{\xi}$. Por lo tanto, la ventana de observación en la que podemos separar diferentes modelos de agrupación, viene dada para valores de $\bar{N}\bar{\xi}$ mayores que la unidad. En la práctica, esto ocurriría en escalas mayores que unos pocos h^{-1} Mpc, donde $\bar{N} \sim R^3$ es grande y domina a $\bar{\xi} \sim R^{-2}$ (Croton et al., 2004b). En escalas más pequeñas, donde $\bar{\xi} > 1$, $\bar{N}\bar{\xi}$ siempre será pequeño, y las muestras de galaxias normalmente serán demasiado escasas para mostrar desviaciones medibles de la contribución gaussiana. Por lo tanto, hay que subrayar que la VPF es un buen discriminante sólo de la agrupación débilmente no lineal. En el régimen altamente no lineal, los vacíos no nos proporcionan mucha información (ver referencias anteriores).

Aunque la expansión dada en la Eq. 3.14 sólo es técnicamente válida para valores pequeños de $\bar{N}\bar{\xi}$, las implicaciones para la agrupación se extienden más allá. Para valores grandes de $\bar{N}\bar{\xi}$, los modelos con diferentes amplitudes jerárquicas S_p dan diferentes RVPF χ : cuando $\bar{N}\bar{\xi}$ aumenta, el valor de χ se reduce y la VPF se hace más grande (con respecto al correspondiente caso de Poisson). La CPDF gaussiana ($S_p = 0$) produce los menores valores de χ y, por tanto, las mayores desviaciones en la VPF. Como ilustraremos con los modelos a continuación, valores mayores de $S_p > 0$ darán lugar a valores mayores de $\xi = \bar{N}\bar{\xi}$.

3.3.3. Modelos fenomenológicos de agrupamiento

Dado que las ecuaciones dinámicas que rigen la agrupación gravitacional no pueden resolverse en el régimen débilmente o fuertemente no lineal utilizando la teoría de perturbaciones, se han propuesto varios modelos fenomenológicos para relacionar las funciones de correlación de orden superior con la función de correlación bipuntual (Croton et al., 2004b, Fry, 1986, Fry et al., 1989, Vogeley et al., 1994). Siguiendo la línea de estos autores compararemos resultados en el Cap. 7 con algunos de estos modelos.

Presentamos a continuación una serie de modelos de relaciones de escala que

difieren en la forma en que fijan los coeficientes de escala S_p . Aquí damos una breve descripción de estos modelos y remitimos a las citas para una descripción más detallada de los mismos.

Modelo mínimo

El primer modelo es el llamado modelo mínimo de agrupamiento, cuyo motivo es considerar una distribución de cúmulos de galaxias aglomeradas, siendo Poisson la propia distribución de cúmulos en el espacio con una ocupación Poisson de galaxias. Esto remite al modelo del halo (e.g. Cooray y Sheth, 2002) pero con un perfil Poisson de halo/cúmulo. La evaluación del conjunto de valores de S_p a partir de la función de distribución generada por este modelo conduce a la siguiente forma funcional para χ :

$$\begin{aligned}\chi &= (1 - e^{-\bar{N}\bar{\xi}})/\bar{N}\bar{\xi} && \text{(mínimo),} \\ S_p &= 1 && \text{(asimetría : } S_3 = 1\text{).}\end{aligned}\tag{3.15}$$

Fry (1986) especuló que este modelo representa un límite inferior de las funciones permitidas $\chi(\bar{N}\bar{\xi})$ en cualquier modelo jerárquico consistente.

Modelo binomial negativo

El segundo modelo, comúnmente llamado modelo binomial negativo, se ha utilizado en varios campos con diferentes motivos físicos (Carruthers y Minh, 1982, Carruthers y Shih, 1983, Elizalde y Gaztanaga, 1992, Fry, 1986, Gaztanaga y Yokoyama, 1993, Klauder et al., 1969). Tras un conjunto T de ensayos independientes con probabilidad q de “éxito” y $p = 1 - q$ de “fracaso”, la probabilidad de tener un número S de éxitos y $F = T - S$ número de fracasos viene dada por la distribución binomial:

$$P(S) = \frac{(F + S)!}{S!F!} (1 - q)^F q^S.$$

La distribución binomial negativa describe la probabilidad de tener un número S de aciertos tras un número fijo F de fallos:

$$P(S) = \frac{(F + S - 1)!}{S!(F - 1)!} (1 - q)^F q^S.$$

Nótese que en el caso binomial lo que se fija es el número total de ensayos. Podemos identificar un “éxito” como el hallazgo de una galaxia en una celda, de modo que $P_N = P(N = S)$ es la CPDF. El número fijo de fallos, F , se supone inversamente proporcional a $\bar{\xi}$ (cuanto mayor sea $\bar{\xi}$, menor será el número de fallos para contar una galaxia en una celda, Croton et al. 2004b). Se asume que la probabilidad de un fallo p es proporcional al producto $\bar{N}\bar{\xi}$ (debido a la agrupación hay un exceso de la raíz cuadrática media de $\bar{N}\bar{\xi}$ de galaxias dentro de una celda con densidad \bar{N} : cuanto mayor sea esta aglomeración, mayor será la probabilidad de no encontrar galaxias en una celda aleatoria). Después de fijar las constantes de proporcionalidad, esto lleva a $F = 1/\bar{\xi}$ y $p = \bar{N}\bar{\xi}/(1 + \bar{N}\bar{\xi})$ (para una derivación diferente ver Gaztanaga y Yokoyama 1993). Este modelo es una versión discreta de la distribución de probabilidad gamma (véase Gaztanaga et al., 2000). La RVPF y los cumulantes en este caso son:

$$\begin{aligned}\chi &= \ln(1 + \bar{N}\bar{\xi})/\bar{N}\bar{\xi} && \text{(binomial negativo),} \\ S_p &= (p - 1)! && \text{(asimetría : } S_3 = 2\text{).}\end{aligned}\tag{3.16}$$

Modelo termodinámico

El tercer modelo fue sugerido por primera vez por Saslaw y Hamilton (1984) y surgió de una teoría termodinámica de las propiedades de la agrupación gravitacional. El modelo original tenía un grado fijo de virialización (varianza de temperatura o de densidad) para todos los tamaños de celdas, pero tal comportamiento es inconsistente con las observaciones. El modelo se amplió posteriormente (véase, por ejemplo, Fry, 1986) para incluir un nivel diferente de virialización en cada escala, que se identifica con la varianza $\bar{\xi}$ en función de la escala. Los resultados son

$$\begin{aligned}\chi &= [(1 + 2\bar{N}\bar{\xi})^{1/2} - 1]/\bar{N}\bar{\xi} && \text{(termodinámico),} \\ S_p &= (2p - 3)!! && \text{(asimetría : } S_3 = 3\text{).}\end{aligned}\tag{3.17}$$

donde $(2p - 3)!! = (2p - 3)(2p - 5)(2p - 7)\dots$ y se trunca en cero.

Hay una relación entre este modelo y el binomial. El modelo termodinámico de agrupación de galaxias predice la probabilidad $f(N)$ de encontrar N galaxias

en un volumen V (Saslaw y Hamilton, 1984),

$$f(N) = \exp[-\bar{N}(1-b) - Nb] \frac{\bar{N}(1-b)}{N!} [\bar{N}(1-b) + \bar{N}b]^{N-1}, \quad (3.18)$$

donde b es el cociente

$$b = \frac{-W}{2K} \quad (3.19)$$

que compara la energía de correlación gravitacional W con la energía cinética K . En el límite virial $b \rightarrow 1$, $\bar{N} \rightarrow \infty$, la probabilidad de vacío reducida para la distribución termodinámica es la Ec. 3.17, y el límite $b \rightarrow \infty$ (que es físicamente imposible en el contexto de la teoría termodinámica) produce la binomial negativa, o la distribución de Bose-Einstein modificada, que tiene una probabilidad de vacío reducida expresada en la Ec. 3.16.

Modelo lognormal

La distribución lognormal (e.g. Coles y Jones, 1991, Weinberg y Cole, 1992) se utiliza a menudo como modelo fenomenológico para la agrupación de galaxias y materia oscura. Aunque no existe una expresión analítica para su RVPF, se puede estimar numéricamente (véanse las referencias anteriores) y se encuentra que se comporta de forma similar al modelo termodinámico (Croton et al., 2004b). Además, estos autores notan que al igual que en el modelo termodinámico, la distribución lognormal también tiene un valor alto para la asimetría: $S_3 = 3 + \bar{\xi}$ (que tiende exactamente al valor termodinámico $S_3 \rightarrow 3$ en grandes escalas donde $\bar{\xi} \rightarrow 0$). Remarcan, finalmente, que hay que tener en cuenta que el modelo lognormal no es verdaderamente jerárquico, ya que no tiene momentos constantes S_p , pero en la práctica las variaciones tienen poco efecto en la RVPF.

Distribuciones poissonianas y gaussianas

Además de los modelos anteriores, también es ilustrativo obtener las expresiones analíticas de la RVPF para distribuciones puramente Poisson y Gaussianas. Trivialmente, a partir de la Eq. 3.13 vemos que

$$\chi = 1 \quad (\text{Poisson}), \quad (3.20)$$

y

$$\begin{aligned}\chi &= 1 - \frac{1}{2}\bar{N}\bar{\xi} && \text{(Guassiano),} \\ S_p &= 0 && \text{(asimetría : } S_3 = 0\text{).}\end{aligned}\tag{3.21}$$

Esto último sólo tiene sentido para valores pequeños de $\bar{N}\bar{\xi}$, pero nótese que incluso cuando la distribución subyacente no es gaussiana, la expresión anterior siempre da una buena aproximación a la probabilidad de vacío en el límite de valores pequeños de $\bar{N}\bar{\xi}$.

El acuerdo entre la RVPF de una dada muestra con alguno de estos modelos de agrupamiento sugiere que los momentos de orden superior de la muestra estudiada podría escalar de forma similar con la función de correlación bipuntual, es decir, las correlaciones de N puntos son consistentes con el escalamiento jerárquico. Por último, uno puede comprobar este comportamiento de escala calculando $\chi(\bar{N}\bar{\xi})$ para muestras con diferentes densidades de galaxias (ver Croton et al., 2004b, y sus referencias). Si las RVPF para todas las muestras colapsan en una única función, entonces los coeficientes S_p son independientes de la escala para $p > 2$ (ya que expresar la VPF como una función de $\bar{N}\bar{\xi}$ elimina la dependencia de la amplitud de la función de correlación bipuntual). En este trabajo nos limitamos a mostrar los modelos mencionados, pero en la literatura pueden encontrarse más (e.g. Croton et al., 2004b, Fry y Colombi, 2013, Mekjian, 2007, y sus referencias).

3.3.4. Resultados previos

Se han realizado numerosos estudios de estadísticas de vacío en diversos relevamientos en *redshift* y simulaciones. A continuación repasamos algunos de los más relevantes que incluyen específicamente la VPF.

Verificación de la relación de escala

Para examinar la dependencia de la VPF con los momentos de orden superior y verificar la relación de escala, Vogeley et al. (1994) traza la RVPF en función de la variable de escala $\bar{N}\bar{\xi}$ usando como datos una combinación de los relevamientos observacionales en el espacio de *redshift* *Center for Astrophysics Redshift Survey*, CfA1 (Huchra et al., 1983) y CfA2 (de Lapparent et al., 1986,

Geller y Huchra, 1989, Huchra, 1989, Huchra et al., 1994), y los compara con tres modelos de agrupamiento: el modelo termodinámico, el negativo binomial y el modelo mínimo (Fig. 3.4). Encuentran que las RVPF para todas las muestras trazan aproximadamente el mismo lugar en el plano $\bar{N}\bar{\xi}$ vs χ que se encuentra entre las curvas de los modelos mínimo y termodinámico. La aparente concordancia de $\chi(\bar{N}\bar{\xi})$ para las diferentes muestras sugiere que los momentos de orden superior de las diferentes muestras podrían escalar de forma similar con la función de correlación bipuntual, es decir, las funciones de p puntos son consistentes con el escalamiento jerárquico. Este resultado implica que, en escalas no lineales, las funciones de correlación de orden superior podrían depender de productos de pares de la función de dos puntos $\xi(r)$. En escalas mayores, sin embargo, encuentran que las VPF de todas las muestras se apartan de la predicción de escalamiento jerárquico. Maurogordato y Lachieze-Rey (1987) también encontraron un escalamiento jerárquico de la VPF en muestras del primer relevamiento CfA (Huchra et al., 1983) y Lachieze-Rey et al. (1992) en muestras del *Southern Sky Redshift Survey* (SSRS da Costa y Pellegrini, 1988).

Un estudio análogo fue realizado posteriormente por Croton et al. (2004b) usando un relevamiento más grande y con mayor densidad de galaxias, el *Two-degree Field Galaxy Redshift Survey* (2dFGRS, Colless et al., 2001), lo cual les permitió medir la distribución de galaxias en un rango más amplio tanto de varianza ($\bar{\xi} \sim 0,3 - 20$) como de número medio de galaxias ($\bar{N} \sim 10^{-4} - 10^2$). Los autores trazan la función de probabilidad de vacío reducida, χ , individualmente como una función tanto del número medio de galaxias, \bar{N} , como de la varianza, $\bar{\xi}$ (Fig. 3.5, paneles superior e inferior respectivamente). La característica principal de esta figura es que ni \bar{N} ni $\bar{\xi}$ muestran individualmente una escala jerárquica cuando se trazan contra χ . Para comprobar la existencia de una escala jerárquica en el 2dFGRS, Croton et al. (2004b) trazan en la Fig. 3.6 la función de probabilidad de vacío reducida, χ , como una función de la variable de escala $\bar{N}\bar{\xi}$. De esta forma se elimina la dependencia de la probabilidad de vacío con la varianza y la densidad media. Si existe un escalamiento entre funciones de correlación de diferentes órdenes, se espera ver que todos los puntos de cada muestra estén sobre la misma curva; efectivamente, la Fig. 3.6 de los autores muestra la clara pauta de un escalamiento jerárquico en los momentos de agrupación del 2dFGRS. Se observa que todos los puntos siguen una trayectoria ajustada (dentro de las barras de error) hasta valores de $\bar{N}\bar{\xi} \sim 30$, y se sitúan cerca de la predicción del

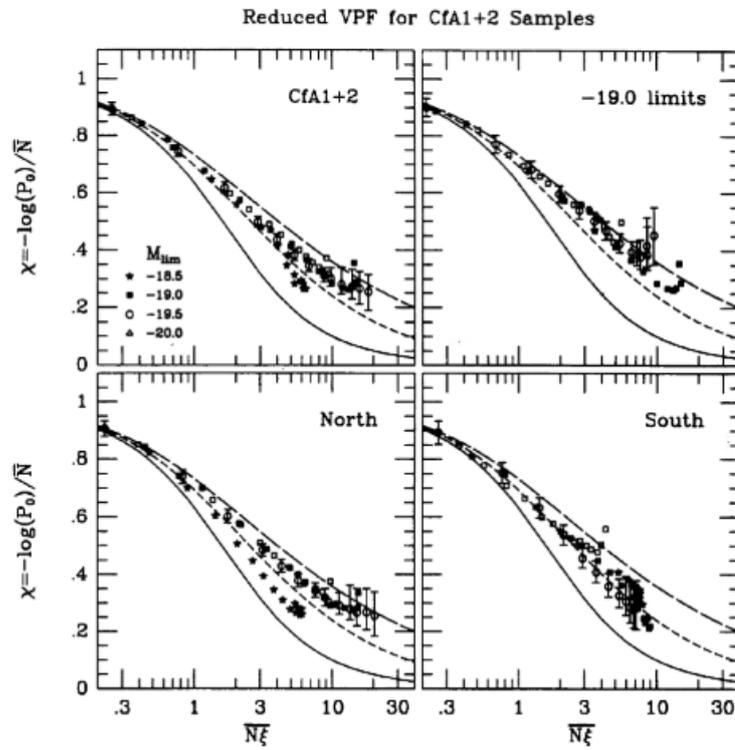


Figura 3.4: La RVPF, $\chi = -\ln P_0/\bar{N}$ como función de la variable de escala $\bar{N}\xi$. Un valor pequeño de χ implica una gran desviación de P_0 para una distribución Poisson con la misma densidad de galaxias. Las curvas representan el modelo termodinámico (trazos largos), el modelo binomial negativo (trazos cortos), y el modelo mínimo (línea sólida). Figura extraída de Vogeley et al. (1994).

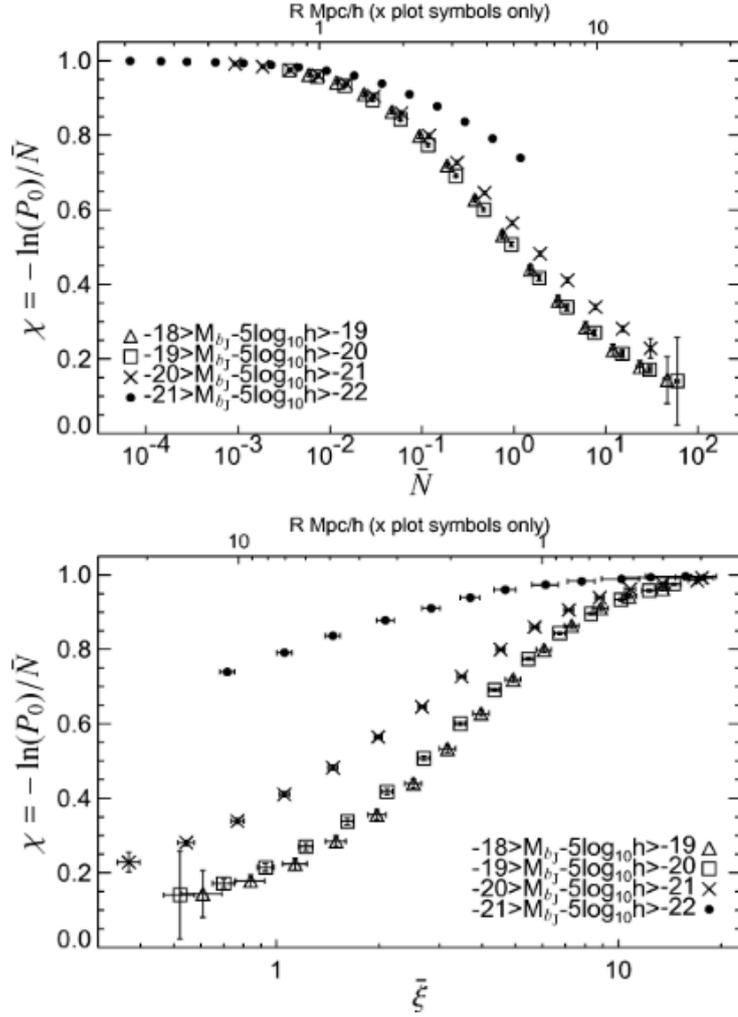


Figura 3.5: La RVPF del relevamiento 2dFGRS, $\chi = -\ln P_0/\bar{N}$, en función del número medio de galaxias \bar{N} (panel superior), y de la varianza de la distribución (panel inferior), $\bar{\xi}$, medida para los catálogos de volumen limitado en distintos intervalos de luminosidad. Los autores notan que ninguna de las dos variables muestra una escala jerárquica cuando se traza individualmente contra ξ . Figura extraída de Croton et al. (2004b).

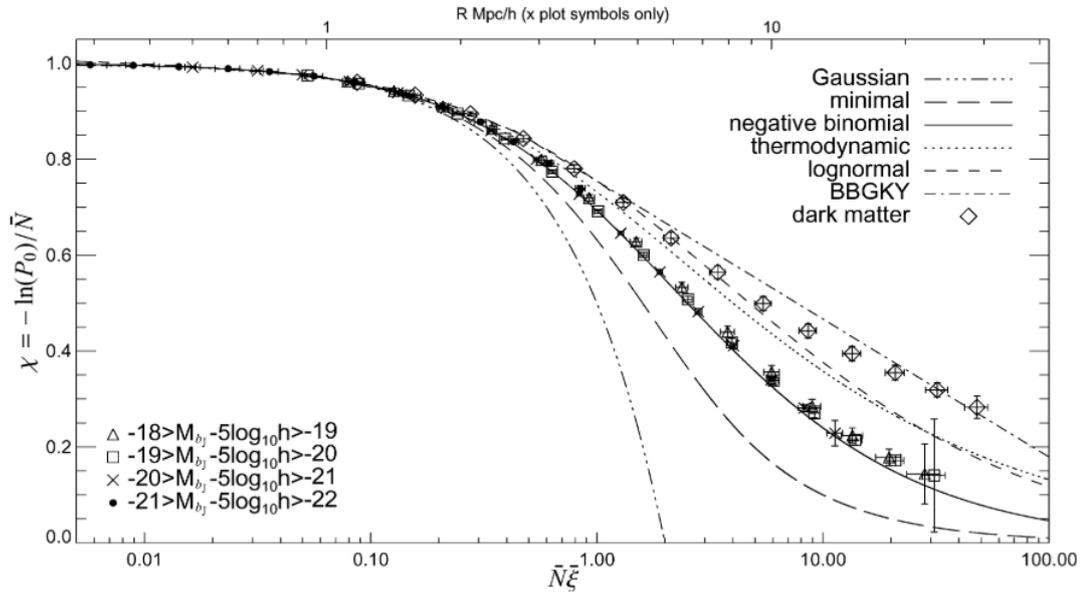


Figura 3.6: La RVPF, $\chi = -\ln P_0/\bar{N}$, en función de la variable de escala $\bar{N}\xi$ para los cuatro catálogos de volumen limitado de galaxias 2dFGRS. La RVPF de la materia oscura, medida a partir de la simulación del volumen de Hubble Λ CDM, se muestra con diamantes grandes. En todos los casos, los valores más pequeños de χ implican mayores desviaciones de una distribución de Poisson. Se ve que todos los puntos colapsan en una sola curva, de modo que la escala jerárquica está presente en la distribución de galaxias. Figura extraída de Croton et al. (2004b).

modelo binomial negativo a lo largo de todo el rango. Los autores argumentan, además, que tales valores abarcan la agrupación de galaxias desde el régimen profundamente no lineal hasta el lineal, revelando un escalamiento jerárquico hasta escalas de $\sim 20h^{-1}\text{Mpc}$ o más. A valores mayores de $\bar{N}\bar{\xi}$ encuentran pequeñas desviaciones del modelo binomial negativo, y remarcan que estas desviaciones son mayores para las muestras de luminosidad más débiles. Relacionan este resultado con la débil dependencia de los coeficientes S_p con la luminosidad de las galaxias encontrada por Croton et al. (2004a), donde se reporta que las muestras más débiles suelen tener valores de S_p mayores que las muestras más brillantes (aunque con grandes barras de error). El efecto de este aumento en el escenario jerárquico daría como resultado un valor de χ más cercano a la unidad (ver Ec. 3.14).

Los autores destacan, en la Fig. 3.6, la inconsistencia de la RVPF con una distribución gaussiana en todas las escalas consideradas (hasta aproximadamente $30h^{-1}\text{Mpc}$ en el 2dFGRS). En las escalas grandes, en las que las funciones de correlación de las galaxias se vuelven demasiado pequeñas para medirlas de forma independiente, encuentran que el valor de \bar{N} aumenta más rápido de lo que disminuye $\bar{\xi}$, y por lo tanto χ sigue estando fuertemente afectada por las correlaciones de orden superior. Esto les lleva a concluir que incluso en el régimen cuasi-lineal, en el que cabría esperar que la agrupación de galaxias fuera sencilla, las correlaciones de orden superior siguen desempeñando un papel importante en la composición de la distribución en gran escala.

Además, Croton et al. (2004b) aplican dos pruebas para ilustrar la robustez de sus resultados y, por lo tanto, el grado de confianza que debería tenerse en su conclusión a favor de la existencia de escalamiento jerárquico en el 2dFGRS. La primera consiste en verificar afirmativamente que el escalamiento se obedece en ambas regiones del 2dFGRS (polo galáctico sur y norte). En la segunda, comprueban las propiedades de escalado calculando la RVPF para muestras de galaxias diluidas al azar. La suposición es que tales diluciones no afectan la función de correlación bipuntual, y dentro del paradigma jerárquico el escalamiento exhibido en la Fig. 3.6 también debería permanecer sin cambios, como se muestra en la Fig. 3.7. Comprueban así que la tendencia de la escala jerárquica existe y sigue el modelo binomial negativo, en consonancia con sus conclusiones anteriores.

En el 2dFGRS, la RVPF parece comportarse de forma algo diferente a la presentada por Vogeley et al. (1994) para las muestras CfA-1 y CfA-2, que muestran más dispersión con magnitud y valores muy por encima del modelo binomial

negativo (comparar figuras 3.4 y 3.6). Croton et al. (2004b) no observan ninguna desviación significativa como la que Vogeley et al. (1994) habían encontrado anteriormente. Mientras que estos encuentran que el escalamiento jerárquico se cumple hasta escalas de $\sim 10h^{-1}\text{Mpc}$, los resultados de Croton et al. (2004b) indican que existe este comportamiento en la distribución de galaxias hasta escalas de al menos $R \sim 20h^{-1}\text{Mpc}$.

El escalamiento jerárquico autode la distribución de galaxias parece verificarse también a distintos *redshift*. Conroy et al. (2005) calcularon la RVPF para $z \sim 1$ usando datos y catálogos simulados basados en el *Deep Extragalactic Evolutionary Probe 2* (DEEP2, Yan et al., 2004) y para $z \sim 0$ con datos del *Sloan Digital Sky Survey - DR2* (SDSS-DR2, Abazajian et al., 2004, York et al., 2000), encontrando que las galaxias reproducen con muy buen acuerdo, y en ambos *redshift*, el modelo de escalamiento binomial negativo. Al igual que Croton et al. (2004b), encuentran que este acuerdo es insensible a diluciones aleatorias de la muestra. Sin embargo, sostienen que esta insensibilidad de la RVPF con respecto a \bar{N} se debe simplemente a la formulación de la VPF como una suma de funciones de correlación de órdenes superiores e insertando el *Ansatz* jerárquico. A su vez, Croton et al. (2006) argumentan que la recuperación de una forma universal para la VPF es justamente una prueba de la escala jerárquica de las funciones de correlación de orden superior de las galaxias.

En cualquier caso, hay un consenso de que la forma de la curva de la VPF de las galaxias está generalmente bien descrita, en el espacio de *redshift*, por el modelo binomial negativo.

Dependencia con luminosidad y color

Con el relevamiento CfA1+2, Vogeley et al. (1994) encuentran que muestras de galaxias con distinta luminosidad tienen un agrupamiento distinto. De las Ecs. 3.12 y 3.14 podemos ver que si las diferencias en la $P_0(R)$ entre distintas muestras de galaxias dependen sólo de la densidad media de la muestra, entonces la RVPF debería ser idéntica para todas las muestras cuando se expresa en función del volumen de prueba, $\chi = \chi(V)$. En la Fig. 3.8 se observa que este escalamiento falla, y por lo tanto concluyen que las galaxias con distinta luminosidad exhiben agrupamiento diferente.

Para complementar esta información con una estadística un poco más conocida, como la función de correlación bipuntual, Vogeley et al. (1994) muestra

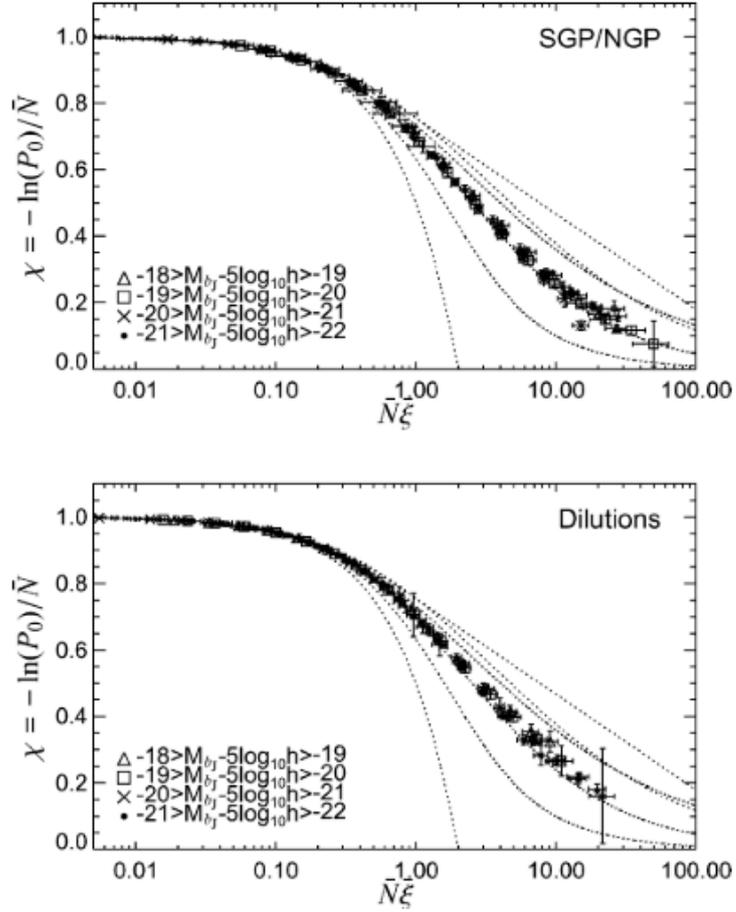


Figura 3.7: Esta figura muestra dos pruebas de confirmación de las propiedades de escalamiento encontradas en el 2dFGRS por Croton et al. (2004b). Los autores muestran en el panel superior que los catálogos de volumen limitado independientes del polo sur (SGP) y polo norte galáctico (NGP) exhiben un escalamiento idéntico al visto en la Fig. 3.6. Aquí los símbolos grandes representan el resultado SGP, y los pequeños el resultado NGP. En el panel inferior se ven los mismos catálogos combinados que en la Fig. 3.6, pero ahora diluidos por factores de 0.5 (símbolos grandes) y 0.25 (símbolos pequeños). Si existe una escala jerárquica en la distribución de galaxias, la dilución debería suponer una pequeña diferencia en los resultados encontrados en la Fig. 3.6. Figura extraída de Croton et al. (2004b).

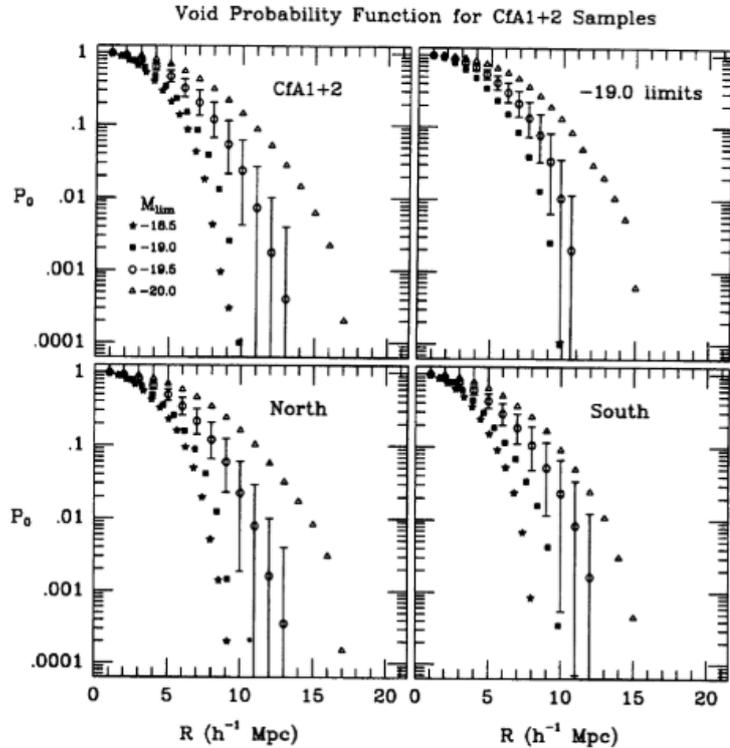


Figura 3.8: Función de probabilidad de vacío $P_0(R)$ para muestras limitadas en volumen del CfA. Las barras de error son la variación de 1σ debido al número finito de volúmenes independientes y a la incertidumbre en la densidad media. Figura extraída de [Vogeley et al. \(1994\)](#).

que la amplitud de dicha función aumenta con la luminosidad de las galaxias y remarca que esta diferencia influye, al menos en parte, a la discrepancia en las curvas de $\chi(V)$ de las distintas muestras. Esta dependencia entre la amplitud de la correlación bipuntual y la luminosidad ya se señalaba en trabajos anteriores (e.g. [Hamilton, 1988](#), [Park et al., 1994](#)) y respaldado por trabajos posteriores ([Zehavi et al., 2011](#)), pero es interesante verlo en el contexto de estadísticos de agrupamiento de más altos ordenes.

Además, [Croton et al. \(2006\)](#) estudiaron la dependencia de la RVPF con la luminosidad y también el color de las galaxias. La Fig. 3.9, extraída de su trabajo, muestra la RVPF para poblaciones de galaxias rojas y azules del 2dFGRS. Aquí, los autores reportan una desviación significativa del escalamiento universal de la VPF reportado en sus trabajos previos: para las galaxias rojas la RVPF se desvía de la escala universal, y tal desviación aumenta para las galaxias más débiles. La RVPF de las galaxias azules, por otro lado, no muestra ninguna dependencia de

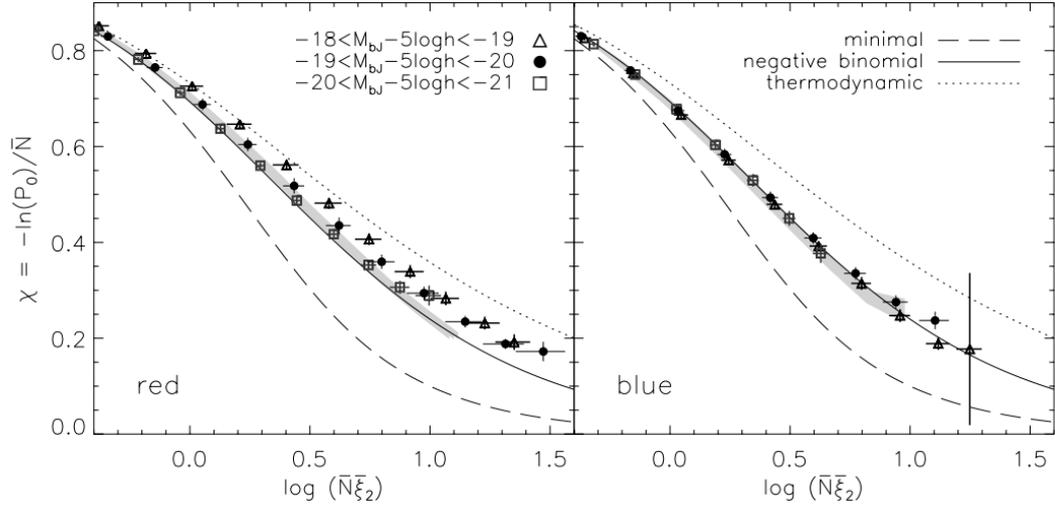


Figura 3.9: La RVPF de galaxias azules y rojas del 2dFGRS trazada en función de la variable de escala $\bar{N}\bar{\xi}$. Los paneles de la izquierda y la derecha muestran las mediciones para las poblaciones roja y azul respectivamente, mientras que los símbolos denotan muestras de galaxias de diferente luminosidad como se indica en la leyenda. Figura extraída de Croton et al. (2006).

la luminosidad, y se ajusta casi perfectamente al modelo binomial negativo en todo el rango representado, como era el consenso general hasta el momento y en acuerdo con sus resultados previos (Croton et al., 2004b). Tinker et al. (2008) llega a esta misma diferencia entre las RVPF de galaxias azules y rojas del *Sloan Digital Sky Survey – Data Release 4* (SDSS–DR4; York et al., 2000).

Espacio real vs. espacio de *redshift*

Es necesario a esta altura remarcar que los modelos jerárquicos de agrupamiento presentados en la Sec. 3.3.3 y usados en la literatura son predicciones para el espacio real. A pesar de esto, hay un acuerdo aparente con los modelos jerárquicos en espacio *real* y la mayoría de los resultados repasados hasta aquí, que han empleado relevamientos en el espacio de *redshift*.

Recordamos de la Sec. 3.2 que en el espacio de *redshift* las velocidades peculiares aumentan la amplitud de las fluctuaciones en gran escala (Kaiser, 1987). Como resultado, los vacíos cósmicos parecen más grandes en este espacio. Por lo tanto, al menos en escalas grandes, es esperable que la VPF, i.e. la probabilidad de encontrar volúmenes sin galaxias, aumente en el espacio de *redshift*. En base a estas ideas, Vogeley et al. (1994) contrastan sus resultados en el espacio de *redshift* con los mismos estadísticos en el espacio real empleando relevamientos

de galaxias simuladas (relevamientos *mock*), y notan que, efectivamente, la P_0 es mayor en escalas grandes en el espacio de *redshift*. Los autores argumentan que en escalas chicas $R < 5h^{-1}\text{Mpc}$, los FOG causados por cúmulos y grupos sobresalen dentro de los vacíos y reducen ligeramente la VPF en el espacio de *redshift* (Vogeley et al., 1991). Este último efecto hace que en escalas chicas la P_0 en *redshift* coincida con la del espacio real porque el efecto Kaiser se compensa con el de los dedos de dios.

Comparando la RVPF de las muestras en el espacio real y en el de *redshift*, Vogeley et al. (1994) encuentran que el escalamiento jerárquico de la VPF en el espacio de *redshift* no implica un escalamiento jerárquico en el espacio real. En la Fig. 3.10 se muestra la RVPF de los relevamientos simulados tanto en el espacio real (conjunto de paneles de la izquierda) como en el espacio de *redshift* (conjunto de paneles de la derecha) para distintos modelos cosmológicos. En el espacio de *redshift*, concuerdan bien con la hipótesis jerárquica, siguiendo uno u otro modelo fenomenológico. Sin embargo, la VPF en el espacio real muestra fuertes desviaciones de la escala jerárquica. Este resultado es quizás poco intuitivo, pues se podría esperar que las velocidades peculiares perturben posibles propiedades de escala en el espacio real. Sin embargo, es un resultado que ha sido también observado por otros autores (e.g. Bernardeau et al., 2002, Conroy et al., 2005, Lahav et al., 1993), y sentaron las bases para dudar de la evidencia propuesta hasta entonces, basada en relevamientos en el espacio de *redshift*, de que el agrupamiento podría ser jerárquico en el espacio real (e.g. Bouchet et al., 1993, Fry et al., 1989, Lachieze-Rey et al., 1992, Maurogordato y Lachieze-Rey, 1987).

Dicho esto, las razones del éxito del modelo binomial negativo para la VPF en el espacio de *redshift* no son claras. Los resultados de Vogeley et al. (1994) parecen indicar que la RVPF de sus simulaciones sólo coinciden con el modelo binomial negativo cuando el patrón de agrupación es distorsionado por movimientos peculiares; mientras que la VPF medida en el espacio real no coincide con el modelo binomial negativo. Sin embargo, los resultados de Croton et al. (2006) discutidos en la subsección anterior implican que los movimientos peculiares de las galaxias no son el principal agente responsable del éxito del modelo binomial negativo. El razonamiento detrás de esta conclusión es que se espera que las galaxias rojas débiles sean principalmente satélites dentro de los cúmulos de galaxias y, por lo tanto, se espera que tengan movimientos peculiares de gran amplitud. Sin embargo, en lugar de ser atraídas hacia el modelo binomial negativo en el espacio χ vs

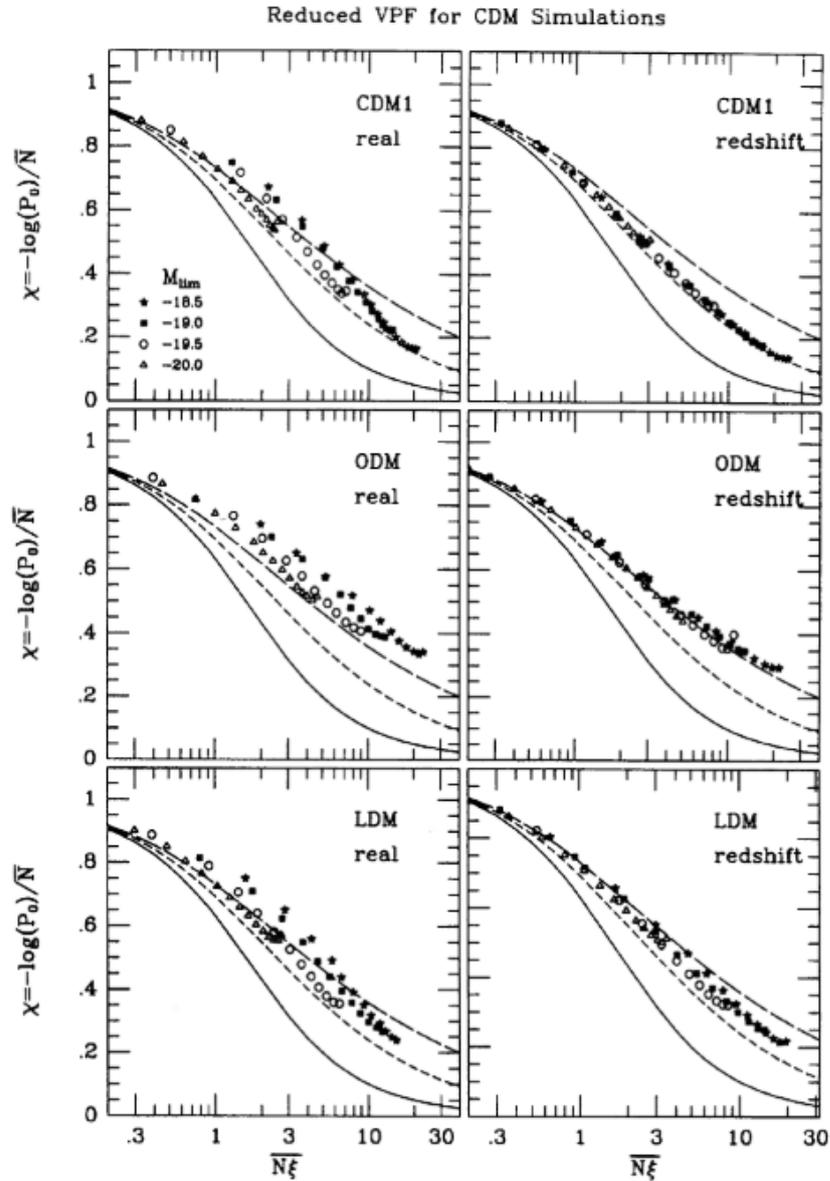


Figura 3.10: RVPF en función de la variable de escala $\bar{N}\xi$ para las simulaciones de los modelos CDM1 (arriba), ODM (centro) y LDM (abajo) (véase el texto para las definiciones) en el espacio real (paneles de la izquierda) y en el espacio de *redshift* (paneles de la derecha). Los tres modelos muestran una sorprendente concordancia con el escalamiento jerárquico en el espacio de *redshift*, pero sus VPF en el espacio real muestran una desviación de este escalamiento. Figura extraída de Vogelely et al. (1994).

$\bar{N}\bar{\xi}$, son las que mayor desviación muestran respecto de este modelo.

Materia oscura y observaciones adicionales

En contraste con las galaxias del 2dFGRS, Croton et al. (2004b) encuentran que la materia oscura sigue una distribución lognormal hasta valores de $\bar{N}\bar{\xi} \sim 6$ (una escala de aproximadamente $R \sim 4h^{-1}\text{Mpc}$ en la simulación), pero luego se desvía fuertemente en escalas mayores (el último punto trazado corresponde a $R = 10h^{-1}\text{Mpc}$ en la simulación; Fig. 3.6). Croton et al. (2004b) argumentan que las diferencias entre las RVPF de galaxias y de materia oscura pueden entenderse observando las diferencias entre sus funciones de correlación promediadas en volumen de orden superior, como muestran Baugh et al. (2004). Con el relevamiento DEEP2, Conroy et al. (2005) encuentran un acuerdo aún más ajustado entre el modelo termodinámico o lognormal y la RVPF de la materia oscura hasta *redshift* $z \sim 2$. En este estudio, además, notan que P_0 es mayor en las galaxias que en la materia oscura para todas las escalas consideradas, indicando que su amplitud de agrupación es mayor y, por lo tanto, que esas galaxias están sesgadas con respecto a la materia oscura. Esto simplemente resulta del hecho de que las galaxias de ese relevamiento simulado habitan halos masivos de materia oscura, que a su vez están más agrupados que la distribución general de materia oscura.

Vogeley et al. (1994) remarcan un fenómeno de apartamiento de los datos respecto a los modelos para los valores más grandes de $\bar{N}\bar{\xi}$ de cada muestra: χ gira hacia arriba y se aleja del locus jerárquico (ver Fig. 3.8). La escala física en la que estos autores detectan el giro es $R \sim 8.5h^{-1}\text{Mpc}$ para las muestras menos luminosas y cerca de $15h^{-1}\text{Mpc}$ para la muestra más luminosa (la correspondencia entre la variable de escala $\bar{N}\bar{\xi}$ y R depende de la densidad numérica del relevamiento). Los autores interpretan este fenómeno como un quiebre en el comportamiento de escala, y por lo tanto concluyen que la distribución de galaxias observada en el espacio de *redshift* sigue el patrón jerárquico en el régimen (débilmente) no lineal, $R < 10h^{-1}\text{Mpc}$, pero se aleja de este comportamiento en escalas mayores.

Conroy et al. (2005) también detecta este fenómeno pero reporta un desvío mucho menor. Estos autores argumentan que puede deberse a las fluctuaciones en las condiciones iniciales de las simulaciones. Además, estas grandes escalas (correspondientes a diámetros de vacío de $\sim 30h^{-1}\text{Mpc}$) se acercan al límite en el que una simulación con un tamaño de caja de $L = 256^3h^{-3}\text{Mpc}^3$ resulta poco fiable debido a que los modos en gran escala, que se vuelven cada vez más im-

portantes a escalas mayores, no pueden incluirse en simulaciones con condiciones de contorno periódicas.

Finalmente, a pesar de que los modelos cosmológicos simulados por Vogelej et al. (1994) son un poco obsoletos actualmente, los resultados que obtienen al calcular la RVPF para estos pueden resultar ilustrativos. En la Fig. 3.10 se muestran resultados de la RVPF para tres modelos cosmológicos: 1) CDM “standard” (CDM1) $\Omega_m = 1$, $h = 0.5$, bias $b = 1.5$ (definido como el cociente entre fluctuaciones de densidad de galaxias y de masa), y $\sigma_8 = 1$ (la raíz cuadrática media de la fluctuación de la densidad de galaxias en una esfera de radio $8h^{-1}\text{Mpc}$), 2) CDM abierto insesgado (ODM) con $\Omega_m = 0.4$, $h = 0.5$, $b = 1$ y $\sigma_8 = 1$, y 3) (LDM) es un modelo CDM con $\Omega_m = 0.4$, $h = 0.6$, $b = 1.3$, $\sigma_8 = 1$ y una constante cosmológica $\lambda = 0.6$ (este modelo es plano y consistente con inflación, quizás el más parecido al modelo estándar ΛCDM actual). La RVPF calculada en espacio de *redshift* del modelo CDM1 se asemeja al modelo binomial, el ODM al termodinámico al igual que la materia oscura en simulaciones (ver referencias mencionadas arriba), lo cual es consistente dado que es el modelo insesgado, y finalmente el modelo LDM presenta mayor dispersión entre las distintas luminosidades ocupando un locus general entre el modelo binomial y termodinámico; a grandes rasgos parece seguir el modelo termodinámico hasta $\bar{N}\bar{\xi} \sim 2-3$. En el espacio real no hay un patrón tan claro entre los distintos modelos cosmológicos.

3.3.5. Resumen y discusión

Una completa y correcta interpretación de la función de probabilidad de vacío, dada su formulación y marco teórico, es complicada de llevar a cabo. En este capítulo apuntamos a reunir información básica necesaria de la literatura para interpretar un estadístico poco usual como es la VPF, y cómo esta se relaciona con órdenes superiores de agrupamiento de galaxias. A continuación repasamos algunos resultados importantes y las discusiones que ellos generan.

¿Es válida la noción de escalamiento jerárquico en la agrupación de galaxias?

Tanto en la teoría de perturbación como en el régimen altamente no lineal, la idea de que la agrupación de orden superior surge de forma jerárquica a partir de la función de correlación bipuntual aparece de forma natural (e.g. Peebles, 1980);

esta “escala jerárquica” sucede a partir de la evolución debida a la inestabilidad gravitacional de una distribución de fluctuaciones de densidad inicialmente gaussiana y tiene evidencia observacional (por ejemplo, Benoist et al. 1999, Bonometto et al. 1993, Bouchet 1993, Fry et al. 1989, Gaztanaga et al. 1992, Maurogordato y Lachieze-Rey 1987; ver Bernardeau et al. 2002 para una revisión).

El *Ansatz* de escalamiento jerárquico fue formalmente derivado asumiendo “agrupamiento estable” y “autosimilaridad” (Bernardeau et al., 2002). La suposición de autosimilaridad es que no hay escalas de tiempo ni de longitud características en la evolución de la estructura en el universo. Esto requiere que $\Omega_m = 1$ e implica que $S_p \sim \xi_p/\xi_2^{p-1}$ es independiente de la escala. Por otro lado, la agrupación estable afirma que, a pequeñas escalas, las regiones de alta densidad se desacoplan de la expansión de Hubble y su tamaño físico es estable, es decir, no cambia con el tiempo (Peebles, 1980). Esto implica que el movimiento relativo de las partículas dentro de las estructuras ligadas gravitacionalmente debería compensar en promedio la expansión de Hubble, y en relación a las funciones de correlación de órdenes superiores implica que los coeficientes S_p son independientes del tiempo. Estas suposiciones, aunque simplistas, han sido útiles a la hora de comprender el comportamiento de las correlaciones a pequeña escala. Por lo tanto, es de esperar que el *Ansatz* de escalamiento jerárquico tenga validez sólo en los regímenes en los que estas aproximaciones tienen una validez razonable.

En el régimen lineal, la teoría de perturbaciones parece validar el escalamiento jerárquico y se ha presentado evidencia observacional consistente con este fenómeno en relevamientos de *redshift* (ver referencias citadas arriba), resultados que repasaremos a continuación.

Un argumento que se ha utilizado para explicar el buen rendimiento de la jerarquía de escalamiento en numerosos trabajos es que las distorsiones del espacio de *redshift* y el sesgo tienden a eliminar la dependencia de los coeficientes con la escala (ver Fig. 3.11). En el espacio de *redshift* las galaxias aparentan tener un agrupamiento mayor debido a los movimientos coherentes de las galaxias al caer hacia el centro de cúmulos, lo cual a su vez aumenta el tamaño de los vacíos. Esto implicaría que, a una dada escala, sea más probable encontrar vacíos en espacio de *redshift* que en el real, y por lo tanto, encontrar un incremento en los valores de S_p (ver Ec. 3.12 o Ec. 3.14). Por otra parte, a escalas pequeñas el crecimiento no lineal se suprime en el espacio de *redshift* debido a la dispersión de la velocidad en cúmulos, y por lo tanto la asimetría y los momentos de orden superior no crecen

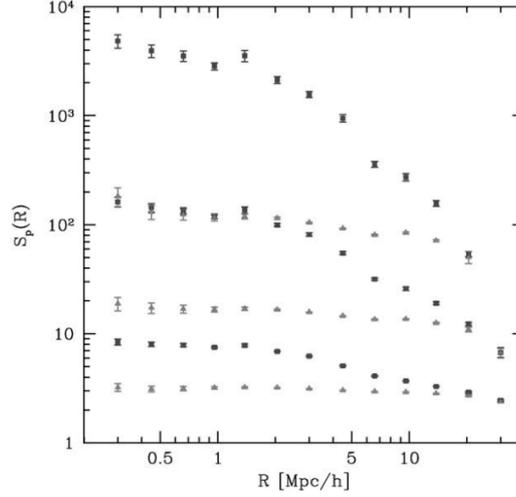


Figura 3.11: Los parámetros S_p para $p = 3, 4, 5$ (de abajo hacia arriba) en espacio de *redshift* para Λ CDM con $\sigma_8 = 0.9$ como función de R . Los cuadrados denotan medidas en simulaciones de N-cuerpos en espacio real, mientras que los triángulos corresponden a valores en espacio de *redshift*, asumiendo la aproximación plano-paralelo. Figura extraída de Bernardeau et al. (2002).

tanto como se esperaría a medida que se sondan escalas más pequeñas (Bonometto et al., 1993, Lahav et al., 1993, Matsubara y Suto, 1994, Scherrer y Gaztanaga, 2001, Suto y Matsubara, 1994). La Fig. 3.11, reproducida de Bernardeau et al. (2002), muestra un ejemplo de los parámetros de S_p en un modelo Λ CDM en el espacio real (cuadrados) y en el espacio de *redshift* (triángulos). Se observa una concordancia entre los dos espacios en las escalas más grandes, y una supresión de los coeficientes en las escalas pequeñas para el caso del espacio de *redshift*, con lo que aparenta ser independiente respecto de la escala. Esto está en acuerdo con los resultados de Vogeley et al. (1994) repasados en la subsección anterior, y con las conclusiones de Lahav et al. (1993), que muestran que las distorsiones del espacio de *redshift* hacen que los momentos de orden superior parezcan casi constantes con la escala. Sin embargo, Croton et al. (2004b) argumentan que los valores de S_p que miden Baugh et al. (2004) y Croton et al. (2004a) no parecen coincidir con las expectativas ni de los modelos de materia oscura ni de los relevamientos de galaxias simuladas (ambos en el espacio de *redshift*). De todas maneras, dada la información presentada, es claro que encontrar evidencia a favor del escalamiento jerárquico en el espacio de *redshift* no implica que haya escalamiento jerárquico en el espacio real.

Modelo binomial negativo: ¿validez universal o coincidencia?

A la luz de la información presentada en este capítulo cabe preguntarse cuáles son los motivos del aparente éxito del modelo jerárquico binomial negativo para representar la distribución de galaxias en el espacio de *redshift*.

Croton et al. (2004b) no reportan ninguna explicación física para el buen rendimiento del modelo binomial negativo, más allá de algunas derivaciones heurísticas (véase la Sec. 3.3.3 y sus referencias). Simplemente señalan que los coeficientes de este modelo tienen valores cercanos a los medidos en el 2dFGRS por Baugh et al. (2004). Por ejemplo, el valor de la asimetría para el modelo binomial negativo es $S_3 = 2$, mientras que la medición directa para el 2dFGRS es $S_3 = 1.86 - 2.03$. A su vez, el modelo termodinámico o la distribución lognormal tienen valores mayores para la asimetría ($S_3 \simeq 3$). Baugh et al. (2004) encuentran una tendencia similar para los coeficientes de orden superior S_4 , S_5 y S_6 . En este sentido, simplemente señalan que si estos coeficientes coinciden con los predichos por un determinado modelo de escalamiento jerárquico, entonces cabría esperar que sus RVPF fueran similares.

Luego de concluir que el modelo binomial negativo es el que mejor predice la forma de la RVPF de muestras de galaxias definidas por luminosidad en el 2dFGRS, Croton et al. (2006) separa la muestra por su color y encuentran que, mientras que las galaxias azules siguen la RVPF predicha por el modelo binomial negativo, las galaxias rojas presentan desviaciones respecto del mismo. Además, el desvío parece ser mayor para las galaxias más débiles en luminosidad. Este resultado está en aparente desacuerdo con interpretaciones previas del éxito del modelo binomial negativo en las que se sugiere que el movimiento peculiar de las galaxias es el agente principal en determinar la forma de la RVPF. Tinker et al. (2008) llega a esta misma discrepancia entre las RVPF de galaxias azules y rojas estudiando estadísticas de vacíos en el contexto de otro método de análisis del agrupamiento de galaxias: la Distribución de Ocupación de Halos (HOD, por sus siglas en inglés; e.g., Jing 1998, Kauffmann et al. 1997, Peacock y Smith 2000).

Además, Tinker et al. (2008) encuentran que las funciones de ocupación que producen funciones de correlación con grandes residuos de una ley de potencia tienden a alejarse del modelo binomial negativo en el espacio χ vs $\bar{N}\bar{\xi}$. La alta fracción de galaxias satélites en la función de ocupación roja produce una fuerte transición del régimen de uno a dos halos que exhiben las galaxias rojas. La función de correlación para la muestra azul es muy cercana a una ley de potencia y,

por tanto, está bien descrita por el modelo de distribución binomial negativo. Esta tendencia también funciona en la dirección opuesta; los HODs que no enfatizan los halos de alta masa se sitúan por debajo de la curva binomial negativa, lo que indica que la binomial negativa no es universal, sino que depende de los detalles de los halos ocupados por una clase determinada de galaxias.

Por último, se ha llegado a conclusiones similares para todo el tratamiento de recuento de celdas. [Hurtado-Gil et al. \(2017\)](#) concluye que el modelo binomial negativo es el mejor ajuste para la distribución de recuento de celdas, i.e. no sólo para P_0 sino también para P_N con $N \geq 1$, mientras que la distribución lognormal (con algunas modificaciones) es una alternativa justificable para celdas que abarquen grandes escalas.

Comentarios finales

Autores como por ejemplo [Conroy et al. \(2005\)](#) y [Tinker et al. \(2006\)](#) sostienen que la VPF queda completamente determinada por la densidad numérica de la muestra y la función de correlación bipuntual promediada en volumen, por lo que no provee necesariamente más información que estas y debería ser usada como estadística complementaria. Recordamos del comienzo de este capítulo, sin embargo, que la VPF es un caso particular de la distribución de recuentos en celdas. Este es un enfoque estadístico simple pero potente de caracterización de la distribución de galaxias en el espacio, ya que contiene información estadística acerca de vacíos y regiones subdensas, cúmulos de diversos tamaños y formas, filamentos, acerca de la probabilidad de encontrar un número arbitrario de vecinos alrededor de posiciones aleatorias, sobre conteos de galaxias en celdas posicionadas aleatoriamente con formas arbitrarias y tamaños aleatorios, y sobre la función de correlación de galaxias de todos los órdenes (ver, e.g. [Yang y Saslaw, 2011](#), y sus referencias). Y a pesar de la gran información que este estadístico contiene acerca del agrupamiento de galaxias, no ha recibido tanta atención como los estadísticos más comunes de agrupamiento tales como la función de correlación bipuntual. La VPF por sí sola puede que sea insuficiente para describir de forma completa la distribución de galaxias, pero es un estadístico complementario que forma parte de este enfoque de recuento de celdas que contiene un gran potencial.

Capítulo 4

Optimizando funciones de correlación: catálogos tipo *glass*

Todos los estimadores requieren un conjunto de puntos que sigan la misma función de selección del relevamiento que se está estudiando, es decir la probabilidad dependiente de la posición de que un objeto esté incluido en la muestra, pero no tienen ninguna agrupación intrínseca. En el espacio de configuración, los estimadores de $\xi(\mathbf{s})$ cuantifican el exceso de probabilidad de encontrar un par de galaxias en un determinado vector de separación \mathbf{s} con respecto a dicha distribución de referencia, a menudo denominada como el “catálogo aleatorio” (Baxter y Rozo, 2013, Davis y Peebles, 1983, Hamilton, 1993b, Landy y Szalay, 1993, Peebles y Hauser, 1974, Vargas-Magaña et al., 2013). La covarianza y el sesgo de estos estimadores dependen del tamaño del catálogo aleatorio, con una muestra de mayor densidad resultando en determinaciones más precisas (Kerscher et al., 2000). Por otro lado, el procesamiento de una muestra aleatoria grande puede ser costoso desde el punto de vista computacional. Por lo tanto, la estimación de $\xi(\mathbf{s})$ suele estar sujeta a un compromiso entre la necesidad de un sesgo y una varianza bajos, y mantener un costo computacional razonable.

Encontrar el equilibrio adecuado entre precisión y costo es especialmente importante en la era de relevamientos galácticos de gran volumen, como el instrumento espectroscópico de energía oscura (DESI, DESI Collaboration et al., 2016), y la misión espacial de la ESA *Euclid* (Laureijs et al., 2011). Este problema se ve agravado por el hecho de que el análisis de los relevamientos de galaxias suele ir acompañado de la medición de las mismas estadísticas de agrupamiento en miles de catálogos simulados que reproducen las propiedades de la muestra real, cada

uno de los cuales debería tener, idealmente, sus propios puntos aleatorios (de Mattia y Ruhlmann-Kleider, 2019). La aplicación de la técnica de reconstrucción de Zeldovich (Eisenstein et al., 2007, Padmanabhan et al., 2012) comúnmente utilizada en los estudios de las oscilaciones acústicas de bariones (BAO, por sus siglas en inglés) representa una complicación adicional, ya que en este caso las estimaciones de $\xi(\mathbf{s})$ requieren calcular los recuentos de pares en dos catálogos aleatorios diferentes.

El estimador más utilizado de $\xi(\mathbf{s})$ es el de Landy y Szalay (1993), en adelante el estimador LS que consiste en contar el número de pares de datos en el catálogo de galaxias observadas con algún vector de separación especificado \mathbf{s} , así como el número correspondiente de pares datos–aleatorios y aleatorio-aleatorio. Como normalmente el catálogo aleatorio es significativamente mayor que la muestra real el componente aleatorio-aleatorio del estimador domina el tiempo total de cálculo. Se han propuesto varias estrategias para reducir el costo computacional de la medición $\xi(\mathbf{s})$, como dividir el catálogo aleatorio en submuestras más pequeñas y promediar los de pares inferidos dentro de cada una de ellas Keihänen et al. (2019). Sin embargo, incluso siguiendo este enfoque el costo computacional de estimar $\xi(\mathbf{s})$ complica el análisis de grandes muestras de galaxias.

Recientemente, Breton y de la Torre (2020) propuso un método para estimar los términos de recuento de pares basado en expresiones analíticas que no depende del uso de un catálogo aleatorio. Este esquema asume que la función de selección del relevamiento puede expresarse como el producto de una huella angular, que se describe mediante mapas pixelados y una distribución radial, que se estima a partir de los recuentos de galaxias. Para los relevamientos cuya función de selección puede ser descrita de esta manera, los resultados obtenidos utilizando este enfoque están en buen acuerdo con los inferidos a partir de los recuentos de pares basados en un catálogo aleatorio. Sin embargo la extensión de este método a las mediciones posteriores a la reconstrucción, donde el campo de desplazamiento inferido a partir de los datos debe aplicarse también al catálogo aleatorio, podría no ser trivial.

Mientras que la varianza de la función de correlación está dominada por la varianza cósmica, en este trabajo nos interesa reducir el error introducido por el componente aleatorio de los estimadores de correlación. El impacto del catálogo aleatorio en la varianza de la función de correlación se debe a sus fluctuaciones de densidad intrínsecas, que se caracterizan por el espectro de potencia $P(k) = 1/\bar{n}$,

donde \bar{n} representa la densidad numérica de los puntos. Aquí evaluamos si los puntos que siguen distribuciones alternativas a la de Poisson, cuyos espectros de potencia exhiben una menor amplitud para el mismo tamaño de muestra, pueden desempeñar el papel del catálogo aleatorio y conducir a una menor varianza en las medidas de agrupación resultantes. Las distribuciones uniformes comúnmente utilizadas para generar condiciones preiniciales en simulaciones de N-cuerpos cubren el volumen de forma más homogénea que una distribución de Poisson, lo que las convierte en un candidato natural para sustituir a los catálogos aleatorios estándar. Algunos ejemplos de estas distribuciones son un *glass* gravitacional (White, 1996) y teselaciones Voronoi con capacidades restringidas (CCVT, por sus siglas en inglés, Balzer et al., 2009, Liao, 2018). Como inconveniente, la generación de esas muestras puede implicar un alto costo computacional.

En este capítulo, basada en el trabajo de Dávila-Kurbán et al. (2021), proponemos utilizar las distribuciones tipo glass obtenidas aplicando iterativamente la misma técnica de reconstrucción utilizada en el contexto de las mediciones de las BAO a un conjunto de puntos inicialmente aleatorios. El pequeño costo computacional adicional que implica la construcción de dicha muestra se ve compensado por la mejora significativa en la varianza de las estimaciones de la función de correlación con respecto a los resultados obtenidos utilizando muestras de Poisson del mismo tamaño. Esto implica también que el mismo nivel de precisión en las mediciones de agrupación puede lograrse con muestras de referencia con densidades significativamente menores, reduciendo así el tiempo total de cálculo.

4.1. Distribuciones puntuales homogéneas e isotrópicas

En esta sección revisamos algunas propiedades relevantes de distribuciones puntuales Poisson y otras distribuciones uniformes y discutimos los algoritmos que se pueden utilizar para generarlas.

Una distribución de Poisson es una distribución estadísticamente homogénea e isotrópica que se caracteriza por un espectro de potencia constante a través de todos los números de onda, $P(k) = 1/\bar{n}$, donde \bar{n} es la densidad del número de puntos. El espectro de potencia determina la varianza normalizada del número de puntos contenidos en esferas de radio R como (véase, por ejemplo, Gabrielli

et al., 2002)

$$\sigma^2(R) = \frac{1}{2\pi^2} \int P(k)W^2(kR)k^2 dk, \quad (4.1)$$

donde $W(kR)$ es la transformada de Fourier de la función de ventana top-hat

$$W(y) = 3 \frac{\text{sen}(y) - y \cos(y)}{y^3}. \quad (4.2)$$

Para una distribución de Poisson, tenemos que $\sigma^2(R) \propto R^{-3}$.

Es posible construir conjuntos de puntos en los cuales $\sigma^2(R)$ disminuye más rápidamente con R que en el caso de Poisson. El decaimiento más rápido posible de cualquier distribución es $\sigma^2(R) \propto R^{-4}$ (Gabrielli et al., 2002). Las muestras que se acercan a este comportamiento límite se denominan a veces *distribuciones de ruido azul* y se describen mediante un espectro de potencia que sigue una ley de potencia $P(k) \propto k^4$, que es la potencia mínima en gran escala esperada para un sistema estocástico discreto (Peebles, 1980). Estas muestras exhiben una potencia significativamente menor que una distribución de Poisson para la misma densidad de puntos.

Uno de los ejemplos que mencionamos de conjuntos de puntos con estas propiedades son las distribuciones tipo *glass* comúnmente utilizadas para establecer las condiciones preiniciales de las simulaciones de N-cuerpos (Baugh y Efstathiou, 1993, Hansen et al., 2007, Joyce et al., 2009, White, 1994). Estas distribuciones se obtienen evolucionando un conjunto de partículas inicialmente distribuidas al azar bajo la acción de una fuerza gravitatoria “negativa” o repulsiva hasta llegar a una configuración de cuasi-equilibrio. En la práctica, generar un *glass* gravitacional de alta calidad con un gran número de partículas es una tarea compleja y los requisitos computacionales asociados son similares a los de una simulación de N-cuerpos.

Se puede obtener una distribución con características similares mediante la construcción de muestras CCVT (Balzer et al., 2009), definidas por la condición de que todos los puntos estén situados en el centro geométrico de sus respectivas celdas de Voronoi, y que todas las celdas tengan aproximadamente el mismo volumen. Estas distribuciones altamente uniformes e isotrópicas se han propuesto recientemente como alternativa al *glass* gravitacional estándar como condiciones pre-iniciales de las simulaciones de N-cuerpos (Liao, 2018). Se generan relajando iterativamente un conjunto de puntos inicialmente aleatorio en una configuración

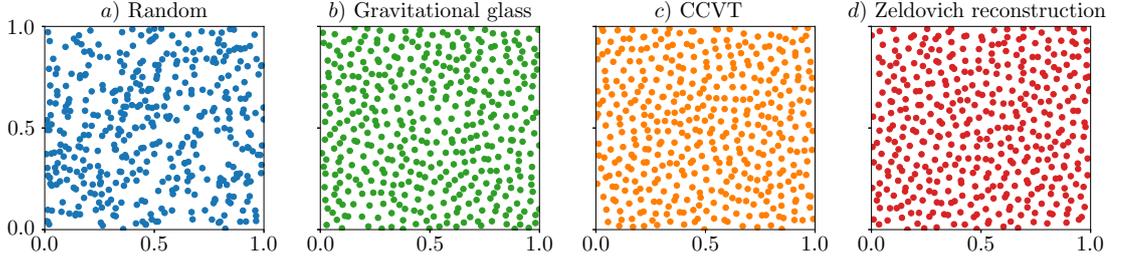


Figura 4.1: Las rodajas que cubren el 10 por ciento de una caja cúbica con 16^3 puntos siguiendo una distribución de Poisson (panel *a*), un glass gravitacional obtenido mediante GADGET-2 (Springel, 2005) (panel *b*), un CCVT construido utilizando el código de Liao (2018) (panel *c*), y una distribución tipo glass obtenida aplicando iterativamente reconstrucción de Zeldovich sobre un conjunto de puntos inicialmente aleatorio.

que satisface las condiciones CCVT (Balzer et al., 2009), pero el costo computacional de este algoritmo se vuelve prohibitivo para un gran número de puntos.

La Fig. 4.1 muestra un corte que cubre el 10 por ciento del ancho de una caja cúbica con 16^3 puntos siguiendo una distribución de Poisson (panel *a*), comparada con un glass gravitacional construido con GADGET-2 (Springel, 2005) y una distribución CCVT obtenida con el código de Liao (2018) con el mismo número de puntos (paneles *b* y *c*, respectivamente). Las distribuciones glass y CCVT cubren el volumen de forma más homogénea que la distribución Poisson, que exhibe mayores fluctuaciones de densidad. Una descripción más cuantitativa de las diferencias entre estas distribuciones puede verse en Fig. 4.2, que muestra los espectros de potencia de las mismas muestras. Mientras que la muestra aleatoria exhibe una potencia constante en todas las escalas, $P(k) = 1/\bar{n}$, las distribuciones glass y CCVT se acercan al espectro de potencia mínimo, $P(k) \propto k^4$, y pasan al comportamiento poissoniano sólo para escalas más pequeñas que la separación media entre partículas, $k \approx 2\pi\bar{n}^{1/3}$. La Fig. 4.3 muestra la varianza $\sigma^2(R)$ derivada de estos espectros de potencia utilizando la Ec. 4.1. Mientras que para la distribución de Poisson $\sigma^2(R) \propto R^{-3}$, la varianza de las distribuciones de ruido azul decae como $\sigma^2(R) \propto R^{-4}$ para escalas mayores que la separación media entre partículas y se aproxima al comportamiento de la muestra aleatoria para escalas menores.

Debido a su menor varianza, un glass gravitacional o un CCVT serían buenos candidatos para reemplazar a las muestras aleatorias estándar utilizadas en la estimación de las estadísticas de agrupación. Sin embargo, generar estas distri-

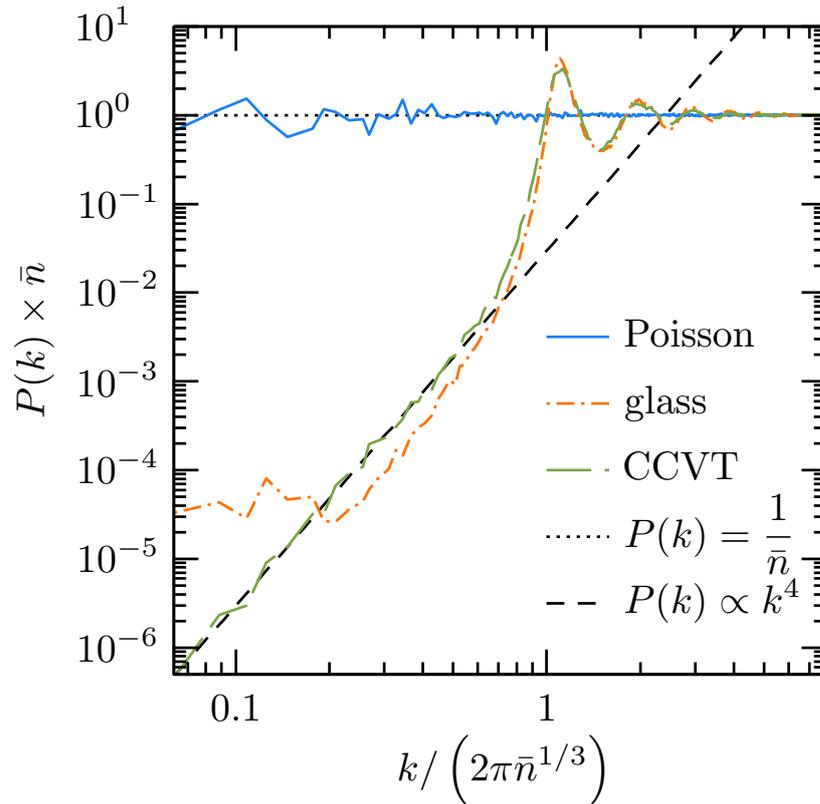


Figura 4.2: Espectro de potencia de las distribuciones de Poisson (sólido), glass (punteado), y CCVT (punteado largo) que se muestran en la Fig. 4.1. Todos los espectros de potencia han sido reescalados por la densidad media de la muestra, \bar{n} , y los números de onda por el equivalente en el espacio de Fourier de la de Fourier de la separación media entre partículas, $2\pi/\bar{n}^{-1/3}$. La muestra de Poisson exhibe una potencia aproximadamente constante en todas las escalas $P(k) = 1/\bar{n}$ (punteada). Las distribuciones glass y CCVT siguen la relación $P(k) \propto k^4$ esperada para distribuciones de ruido azul (punteado).

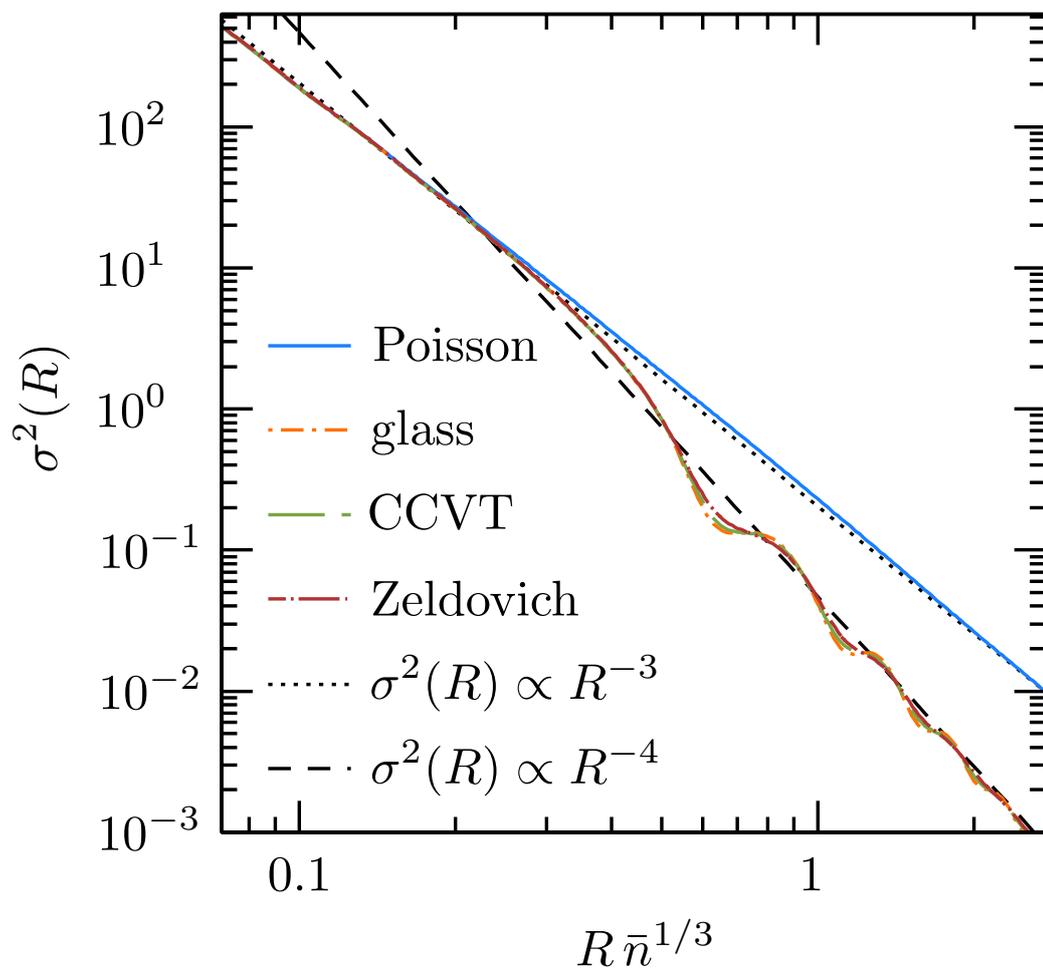


Figura 4.3: Varianza normalizada del número de puntos contenidos en esferas de radio R correspondientes a las distribuciones mostradas en la Fig. 4.1, derivada de sus espectros de potencia utilizando la Ec. 4.1. Mientras que para la muestra de Poisson $\sigma^2(R) \propto R^{-3}$ en todas las escalas, para las distribuciones de ruido azul la varianza decae como $\sigma^2(R) \propto R^{-4}$ para escalas mayores que la separación media entre partículas y se convierte en el comportamiento de la muestra aleatoria para escalas más pequeñas.

buciones con el número de puntos necesarios para esta tarea sería costoso tanto en tiempo como en recursos informáticos. Aunque las condiciones pre–iniciales de las simulaciones de N–cuerpos se construyen a menudo mediante mosaicos de pequeñas distribuciones periódicas de tipo glass para reducir su costo informático, la periodicidad resultante de la distribución podría tener efectos no deseados en la estimación de los recuentos de pares. Como veremos en la Sección 4.2, es posible construir grandes conjuntos de puntos con propiedades similares siguiendo un procedimiento mucho más sencillo.

4.2. Reconstrucción de Zeldovich

La teoría de perturbación lagrangiana ofrece una descripción precisa de la dinámica gravitacional para pequeñas fluctuaciones de densidad. En esta sección, mostramos que también puede utilizarse como alternativa a un cálculo completo de fuerzas de un sistema de N–cuerpos para evolucionar una muestra de puntos inicialmente aleatoria hacia un estado similar al de catálogos glass bajo fuerzas gravitacionales repulsivas con un costo computacional significativamente menor.

La cantidad clave en la teoría de perturbación lagrangiana es el campo de desplazamiento $\Psi(\mathbf{q}, t)$, que mapea la posición inicial (lagrangiana) de un elemento de fluido, \mathbf{q} , a su contraparte euleriana en un momento dado, $\mathbf{x}(\mathbf{q}, t)$, como

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + \Psi(\mathbf{q}, t)^1. \quad (4.3)$$

Se puede encontrar una solución para el campo de desplazamiento imponiendo conservación de la masa entre los sistemas de coordenadas lagrangiano y euleriano, es decir

$$\bar{\rho} d^3q = \rho(\mathbf{x}, t) d^3x, \quad (4.4)$$

donde $\bar{\rho}$ es la densidad media y $\rho(\mathbf{x}, t) = \bar{\rho}(1 + \delta(\mathbf{x}, t))$ representa la densidad euleriana en la posición \mathbf{x} y el tiempo t . Así, manteniendo sólo los términos a nivel lineal, el campo de desplazamiento está relacionado con las fluctuaciones de densidad en el espacio euleriano por

$$\nabla_{\mathbf{q}} \cdot \Psi_{(1)}(\mathbf{q}, t) = -\delta_{(1)}(\mathbf{x}, t). \quad (4.5)$$

¹Esta ecuación es equivalente a la Ec. 1.49.

El subíndice (1) indica que se trata de términos de primer orden. Suponiendo que Ψ es un campo vectorial irrotacional, la solución de esta ecuación se puede escribir en el espacio de Fourier como

$$\Psi_{(1)}(\mathbf{k}) = -\frac{i\mathbf{k}}{k^2}\delta_{(1)}(\mathbf{k}, t). \quad (4.6)$$

Esta es la solución para el campo de desplazamiento en la teoría de perturbación lagrangiana de primer orden y corresponde a la aproximación estándar de Zeldovich (Zel'dovich, 1970).

La Ec. 4.6 puede generalizarse para tener en cuenta las distorsiones lineales del espacio de *redshift* y el sesgo de las galaxias y es la base de la técnica de reconstrucción de Zeldovich que se aplica comúnmente al análisis de galaxias para mejorar la detección de las BAO (Burden et al., 2015, Eisenstein et al., 2007, Padmanabhan et al., 2012). La aplicación del campo de desplazamiento de la Ec. 4.6 con el signo contrario puede deshacer parcialmente los efectos de la evolución gravitatoria no lineal. El mismo principio básico puede aplicarse a un conjunto de puntos aleatorios para suavizar sus fluctuaciones de densidad en gran escala.

La Fig. 4.4 muestra cómo la aplicación iterativa de la reconstrucción de Zeldovich modifica el espectro de potencia del mismo conjunto de puntos aleatorios mostrados en el panel *a*) de la Fig. 4.1. Adaptamos el código de reconstrucción disponible públicamente de Bautista et al. (2018)², que se basa en el algoritmo del espacio de Fourier de Burden et al. (2015). Utilizamos una transformada de Fourier rápida con una resolución de malla tal que el tamaño de la celda viene dado por una cuarta parte de la separación media entre partículas. A continuación, estimamos las densidades utilizando un kernel gaussiano con una escala de suavizado de dos veces el tamaño de la celda. Partiendo de un valor constante en todas las escalas, la amplitud de $P(k)$ disminuye hacia el comportamiento deseado del ruido azul, convergiendo en la potencia mínima $P(k) \propto k^4$ para escalas mayores que la separación media entre partículas. La curva roja discontinua y punteada en la Fig. 4.3 muestra la varianza normalizada $\sigma^2(R)$ correspondiente a la distribución de puntos obtenida tras 50 iteraciones, que decae con la mayor pendiente posible, $\sigma^2(R) \propto R^{-4}$, para escalas mayores que la separación media entre partículas. El panel *d*) de la Fig. 4.1 muestra un corte de esta distribución final de puntos. Los puntos cubren el volumen de la caja de forma mucho más

²github.com/julianbautista/eboss_clustering

uniforme que en el caso Poisson. El procedimiento de aplicar la Ec. 4.6 a una distribución de puntos de Poisson es significativamente más simple y menos costoso computacionalmente, que generar un glass con un código completo de N-cuerpos o construir una distribución CCVT. Este procedimiento también puede aplicarse a grandes muestras de puntos.

4.3. Modificación del estimador de Landy-Szalay

El estimador más utilizado de la función de correlación bipuntual es el de Landy y Szalay (1993), presentado en la Sec. 3.1 con la Ec. 3.4. Aquí lo reescribimos en términos del vector distancia \mathbf{s} en el espacio de *redshift*, y haciendo implícitas las normalizaciones, como

$$\xi_{\text{LS}}(\mathbf{s}) = \frac{DD(\mathbf{s}) - 2DR(\mathbf{s}) + RR(\mathbf{s})}{RR(\mathbf{s})}, \quad (4.7)$$

donde tenemos, análogamente a la Ec. 3.4, recuentos de pares de datos ($DD(\mathbf{s})$) y de posiciones aleatorias ($RR(\mathbf{s})$), y de pares formados por un elemento del conjunto de datos y uno del catálogo de posiciones aleatorias ($DR(\mathbf{s})$) para un vector de separación \mathbf{s} en el espacio de *redshift*, normalizado al número total de pares en cada caso.

El estimador LS de la Ec. 4.7 proporciona la varianza mínima cuando $|\xi| \ll 1$ y es insesgado en el límite de un número infinito de puntos aleatorios, $N_r \rightarrow \infty$. La práctica estándar para lograr una alta precisión en las mediciones de $\xi(\mathbf{s})$ es utilizar un número de puntos aleatorios, N_r , mucho mayor que el tamaño del catálogo de datos real, N_d , a menudo caracterizado en términos de la relación $\alpha \equiv N_r/N_d$. Sin embargo, hay que tener en cuenta que la cantidad más relevante para controlar el sesgo y la varianza del estimador LS es la densidad numérica del catálogo aleatorio y no el valor de α .

Un mayor N_r implica un aumento del número de recuentos de pares y, por tanto, del costo computacional total. Como se discute en Keihänen et al. (2019), mientras que la estimación de la $RR(\mathbf{s})$ domina el tiempo total de cálculo, el error del estimador está dominado por el término $DR(\mathbf{s})$. Como forma de acelerar la estimación de $\xi(\mathbf{s})$ sin que disminuya su precisión, Keihänen et al. (2019) propuso dividir el conjunto total de puntos aleatorios en M_s submuestras y aproximar el total $RR(\mathbf{s})$ por el promedio de los recuentos aleatorios normalizados inferidos

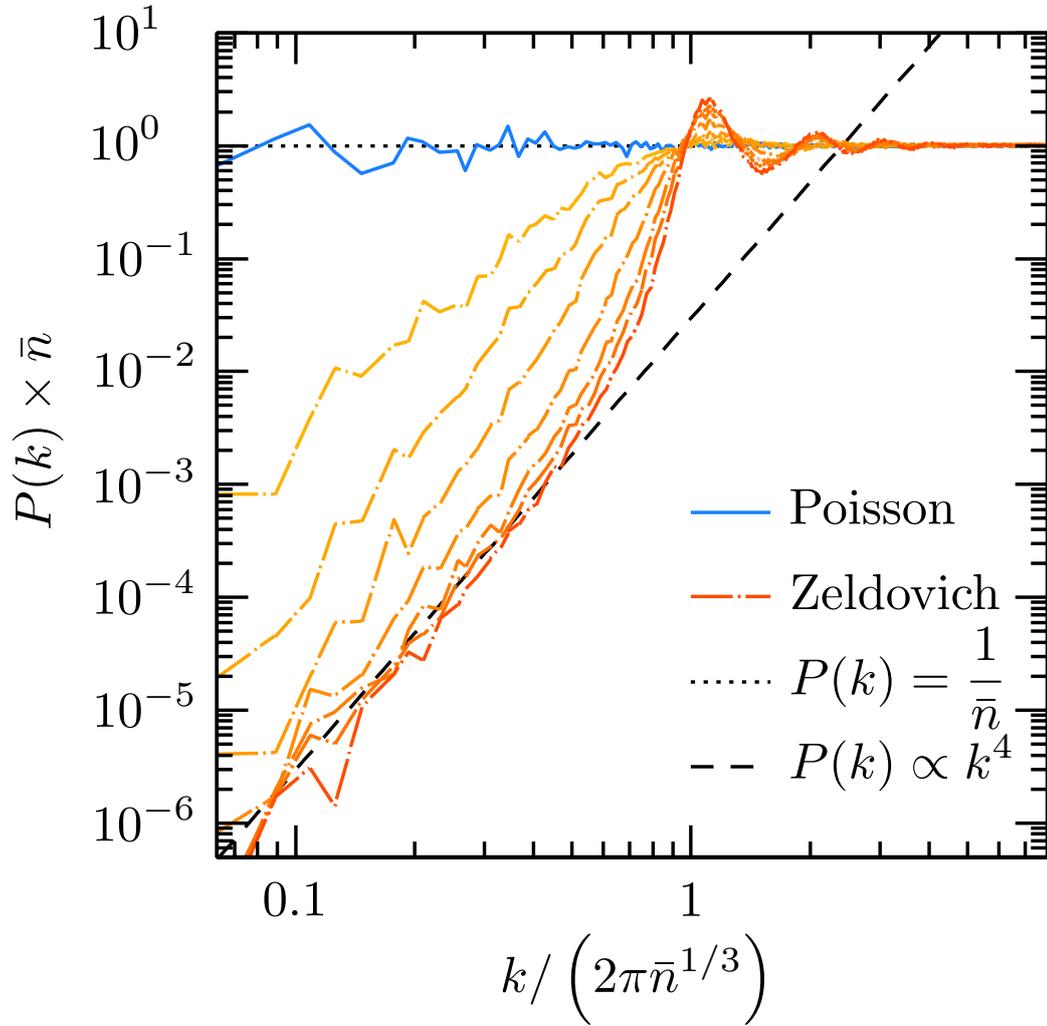


Figura 4.4: Espectros de potencia de las muestras obtenidas aplicando iterativamente la reconstrucción de Zeldovich a un conjunto de puntos aleatorios (líneas discontinuas), mostrados en las mismas unidades que en la Fig. 4.2. Los resultados mostrados corresponden a la configuración inicial (líneas sólidas) y a las iteraciones 1, 2, 4, 10, 20, 30 y 50. El espectro de potencia evoluciona desde el ruido de disparo puro $P(k) = 1/\bar{n}$ (líneas punteadas) hasta la forma mínima forma $P(k) \propto k^4$ (líneas discontinuas) para escalas mayores que la separación media entre partículas.

dentro de cada uno de estos subconjuntos, es decir

$$RR(\mathbf{s}) \simeq \frac{1}{M_s} \sum_{i=1}^{M_s} RR_i(\mathbf{s}), \quad (4.8)$$

donde $RR_i(\mathbf{s})$ representa los resultados inferidos de la i -ésima submuestra aleatoria. Este enfoque, denominado “split-random”, puede reducir el tiempo de cálculo, sin afectar a la varianza o al sesgo del estimador. Sin embargo, si el análisis debe realizarse también en un gran número de catálogos simulados, un gran N_r seguiría implicando un elevado costo computacional.

Proponemos seguir un enfoque alternativo para reducir el sesgo y la varianza de las estimaciones de $\xi(\mathbf{s})$ abandonando el uso de puntos aleatorios, empleando en su lugar distribuciones más uniformes como las descritas en la Sec. 4.1. El uso de distribuciones de puntos tipo glass reduciría el espectro de potencia $P(k)$ de la muestra de referencia en escalas más grandes, lo que resulta en una menor varianza de los recuentos de pares para el mismo número de puntos sin aumentar el tiempo de cálculo. En particular, proponemos utilizar las muestras de puntos de tipo glass obtenidas tras la aplicación iterativa de la reconstrucción de Zeldovich a una distribución inicialmente aleatoria como se discute en la Sec. 4.4.

La simple sustitución de la muestra aleatoria por un glass en el estimador LS conduciría a medidas de agrupación sesgadas. A medida que la muestra se desvía de una distribución de Poisson, la posición de los puntos se vuelve correlacionada y produciría una estimación sesgada de $RR(\mathbf{s})$. Este problema puede evitarse utilizando dos distribuciones de puntos independientes tipo glass o *catálogos glass*, G_1 y G_2 , y modificando el estimador de Landy-Szalay como

$$\xi_{LS,G}(\mathbf{s}) = \frac{DD(\mathbf{s}) - DG_1(\mathbf{s}) - DG_2(\mathbf{s}) + G_1G_2(\mathbf{s})}{G_1G_2(\mathbf{s})}. \quad (4.9)$$

Exceptuando $DD(\mathbf{s})$, todos los términos que aparecen en esta expresión representan recuentos de pares cruzados entre diferentes muestras. De hecho, la Ec. 4.9 se parece a la generalización del estimador de Landy-Szalay comúnmente utilizado para calcular la función de correlación cruzada entre dos datos diferentes (e.g. Blake et al., 2006). La Sec. 4.4 muestra generalizaciones de otros estimadores comúnmente utilizados de las funciones de correlación de dos y tres puntos para el uso de catálogos glass.

Aunque la notación de la Ec. 4.9 está pensada para especificar el uso de las distribuciones glass también puede implementarse con dos catálogos aleatorios

distintos, R_1 y R_2 , es decir

$$\xi_{2R}(\mathbf{s}) = \frac{DD(\mathbf{s}) - DR_1(\mathbf{s}) - DR_2(\mathbf{s}) + R_1R_2(\mathbf{s})}{R_1R_2(\mathbf{s})}. \quad (4.10)$$

La comparación del sesgo y la varianza de los resultados obtenidos mediante este estimador y la Ec. 4.9 puede utilizarse para evaluar el impacto de la sustitución de las muestras aleatorias estándar por catálogos glass del mismo tamaño. La Ec. 4.10 es similar al método de “split-random” de la Ec. 4.8 con $M_s = 2$ pero utilizando los pares cruzados entre las dos submuestras para inferir $RR(\mathbf{s})$. Como métrica de referencia de estos diferentes casos, utilizamos los resultados del estimador LS estándar con un catálogo aleatorio único que contiene el doble de puntos.

4.4. Rendimiento del estimador LS modificado

Metodología

Las condiciones de contorno periódicas nos permiten estimar la función de correlación bipuntual de muestras extraídas de simulaciones de N-cuerpos sin el uso de un catálogo aleatorio como

$$\xi(\mathbf{s}) = DD(\mathbf{s}) \frac{V}{\delta V(\mathbf{s})} - 1, \quad (4.11)$$

donde V representa el volumen de la simulación y $\delta V(\mathbf{s})$ es el volumen del intervalo centrado en la separación de pares \mathbf{s} utilizada para los recuentos de pares (por ejemplo, el volumen de la cáscara esférica entre los radios s y $s + \delta s$ para la función de correlación promediada en ángulo). La comparación de esta medida exacta con los resultados de los estimadores descritos en la Sec. 4.3 con catálogos aleatorios y glass de diferentes tamaños puede utilizarse para cuantificar su varianza y su posible sesgo.

Para ello utilizamos una realización de la suite de simulación de N-cuerpos *Minerva* (Grieb et al., 2016, Lippich et al., 2019). Estas simulaciones siguen la evolución del campo de densidad de la materia oscura sobre una caja cúbica de longitud de lado $L_{\text{box}} = 1.5 h^{-1} \text{Gpc}$ con 1000^3 partículas. Cada simulación representa una realización de un modelo plano ΛCDM con materia oscura física y densidades de bariones $\omega_c = 0.1154$ y $\omega_b = 0.02224$, un parámetro adimensional

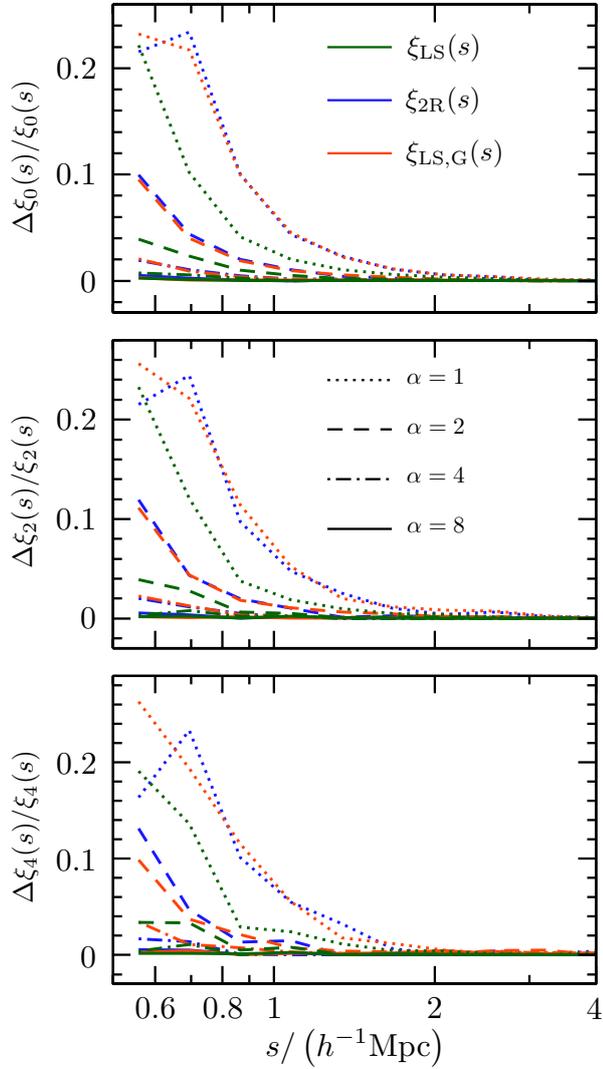


Figura 4.5: Sesgo relativo $\Delta\xi_\ell/\xi_\ell(s)$ de las estimaciones del monopolo (panel superior), cuadrupolo (panel central) y hexadecapolo (panel inferior). Los colores de las líneas indican los resultados de los diferentes estimadores descritos en la Sec. 4.3, mientras que los estilos de las líneas representan el número total de puntos en los catálogos aleatorio y glass, caracterizados por valores de $\alpha = 1, 2, 4$ y 8 . El estimador de la Ec. 4.9 basado en los catálogos glass muestra un rendimiento similar que los resultados obtenidos utilizando muestras aleatorias del mismo tamaño sin introducir ningún sesgo adicional.

de Hubble $h = 0.695$, un índice espectral escalar $n_s = 0.968$, y una amplitud de las fluctuaciones de densidad caracterizada por una raíz cuadrática media de la fluctuación de masa de teoría lineal en esferas con radio de 12 Mpc, $\sigma_{12} = 0,805$ (Sanchez, 2020). Los halos y subhalos de estas simulaciones en *redshift* $z = 0,57$ se poblaron de acuerdo con una distribución de ocupación del halo (HOD) diseñada para generar galaxias sintéticas con propiedades de agrupación comparables a la muestra del CMASS del *Baryon Oscillation Spectroscopic Survey* (Dawson et al., 2013, Reid et al., 2016). Estos catálogos HOD tienen una densidad numérica media de $\bar{n}_d \approx 4 \times 10^{-4} h^3 \text{Mpc}^{-3}$. El impacto de las distorsiones del espacio de *redshift* en las posiciones de las galaxias se añadió teniendo en cuenta la componente de sus velocidades peculiares a lo largo de un eje cartesiano de la caja, situando así un observador en el infinito.

La función de correlación anisotrópica completa puede describirse en términos del módulo de la separación de pares $s = |\mathbf{s}|$ y el coseno del ángulo entre el vector de separación \mathbf{s} y la línea de visión, μ . La información de $\xi(s, \mu)$ puede descomponerse en multipolos de Legendre $\xi_\ell(s)$, dados por

$$\xi_\ell(s) = \frac{2\ell + 1}{2} \int_{-1}^1 L_\ell(\mu) \xi(\mu, s) d\mu, \quad (4.12)$$

donde $L_\ell(\mu)$ denota el polinomio de Legendre de orden ℓ . En primer lugar, utilizamos los estimadores descritos en la Sec. 4.3 para medir $\xi(s, \mu)$ y luego utilizamos los resultados para calcular los multipolos de Legendre con $\ell = 0, 2, 4$. Consideramos dos configuraciones: una centrada en separaciones de pares grandes e intermedias con 26 intervalos lineales en el rango $20 h^{-1}\text{Mpc} \leq s \leq 150 h^{-1}\text{Mpc}$, y la segunda centrada en escalas pequeñas con 14 intervalos logarítmicos para separaciones de pares $0.5 h^{-1}\text{Mpc} \leq s \leq 40 h^{-1}\text{Mpc}$.

Consideramos catálogos aleatorios con densidades numéricas medias \bar{n} correspondientes a 1, 2, 4 y 8 veces la de la muestra HOD de Minerva. Para cada caso, generamos 500 catálogos aleatorios y utilizamos el estimador estándar LS para obtener el mismo número de estimaciones independientes de los multipolos de Legendre $\xi_{\ell=0,2,4}(s)$ de la muestra HOD de Minerva. A continuación, dividimos cada uno de estos catálogos aleatorios en dos subconjuntos con igual número de puntos y los utilizamos para medir los mismos multipolos aplicando el estimador de la Ec. 4.10 y el método de split-random de la Ec. 4.8 con $M_s = 2$. Estos sub-catálogos se tomaron entonces como puntos iniciales para la aplicación iterativa de la reconstrucción de Zeldovich, como se describe en la Sec. 4.4 para generar

catálogos glass independientes que se utilizaron en el estimador de la Ec. 4.9. Las estimaciones de los multipolos obtenidos en cada caso se utilizaron para evaluar el rendimiento de los distintos estimadores. La dispersión de los resultados recuperados utilizando diferentes catálogos aleatorios, $\sigma_{\xi_\ell}(s)$, cuantifica la varianza de estos estimadores, y el valor medio de sus diferencias con el resultado exacto de la Ec. 4.11, $\Delta\xi_\ell(s)$, es una medida de su sesgo.

Sesgo y varianza de los multipolos $\xi_\ell(s)$

Primero nos centramos en el sesgo de los estimadores discutidos en la Sec. 4.3 estudiando las diferencias medias entre los multipolos de Legendre obtenidos utilizando diferentes catálogos aleatorios y catálogos glass y el resultado exacto de la Ec. 4.11. La Fig. 4.5 muestra el sesgo medio relativo $\Delta\xi_\ell/\xi_\ell(s)$ de los estimadores, representados por diferentes colores. Los estilos de línea indican el número total de puntos en los catálogos aleatorios y catálogos glass. Los paneles muestran por separado los resultados para los multipolos $\ell = 0, 2, 4$. Sólo mostramos resultados en separaciones de pares pequeñas, ya que ninguno de los estimadores introduce un sesgo significativo en escalas mayores. En escalas significativamente menores que la separación media entre partículas de los catálogos aleatorios y glass, el estimador de la Ec. 4.9 basado en los catálogos glass muestra un rendimiento similar al de los resultados obtenidos utilizando muestras aleatorias del mismo tamaño.

Las varianzas de los multipolos obtenidas en los diferentes casos en escalas grandes e intermedias se muestran en el conjunto izquierdo de paneles de la Fig. 4.6 mientras que las de escalas pequeñas en el conjunto de paneles derecho. Para juzgar mejor el rendimiento de los nuevos estimadores de las Ecs. (4.9) y (4.10) mostramos los resultados reescalados por la varianza del caso estándar de LS utilizando una muestra aleatoria con el mismo número total de puntos. Al igual que en la Fig. 4.5, los diferentes estilos de línea indican el número de puntos en los catálogos aleatorio y glass. En escalas grandes, el estimador de la Ec. 4.10 (líneas azules) tiene esencialmente el mismo rendimiento que para el caso estándar de LS. Hemos comprobado que el método split-random con $M_s = 2$ da un rendimiento similar. Ambos métodos son, pues, alternativas viables al estimador estándar LS en gran escala para reducir el tiempo total de cálculo del término RR sin sacrificar la precisión de las mediciones. Sin embargo, en escalas más pequeñas que la separación media entre partículas, las varianzas de estos

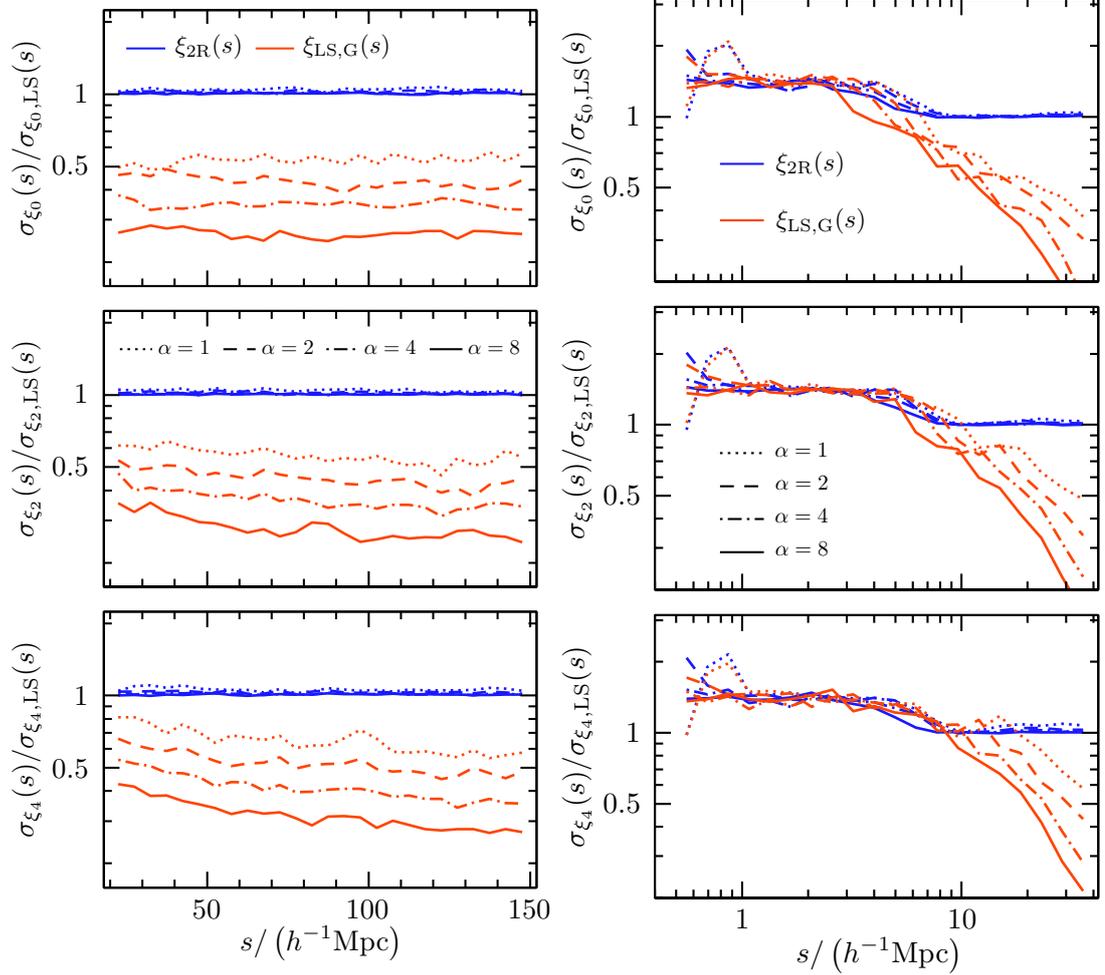


Figura 4.6: *Izq.* La desviación estándar en gran escala de los multipolos de Legendre $\xi_\ell(s)$ con $\ell = 0, 2, 4$ recuperada utilizando los estimadores de las Ecs. (4.9) (rojo) y (4.10) (azul), en relación con la del caso LS basado en el mismo número total de puntos aleatorios. Los diferentes estilos de línea indican el número de puntos de los catálogos aleatorio y catálogos glass, correspondientes a $\alpha = 1, 2, 4$ y 8 . Mientras que el estimador $\xi_{LS,2R}(s)$ tiene un rendimiento similar al LS estándar, la varianza obtenida utilizando catálogos glass es significativamente menor. *Der.* Desviación estándar de los multipolos en escalas más pequeñas o del orden de la media separación entre partículas de las muestras aleatorias y catálogos glass. En este régimen, los estimadores de las Ecs. (4.9) (rojo) y (4.10) (azul) muestran un rendimiento similar y resultado un $\sigma_{x_{i_\ell}}(s)$ que es ~ 30 por ciento mayor que el del resultado estándar de LS basado en el mismo número total de puntos.

estimadores aumentan hasta un nivel de aproximadamente un 30 por ciento más alto que el del resultado completo de LS.

La sustitución de los catálogos aleatorios por muestras similares a las del catálogo glass conduce a una dependencia de escala notablemente diferente de σ_{ξ_ℓ} que el caso estándar de LS. Para escalas menores que la separación media entre partículas, la varianza del estimador de la Ec. 4.9 coincide con la obtenida utilizando dos catálogos aleatorios. Esto es de esperar ya que, como se discute en la Sec. 4.1, en escalas pequeñas el espectro de potencia de las muestras se asemeja al de una distribución de Poisson del mismo tamaño. Sin embargo, en escalas grandes la varianza de $\xi_{\text{LS,G}}(\mathbf{s})$ es significativamente menor que la del estimador LS utilizando una muestra aleatoria con el mismo número total de puntos. La mejora con respecto a LS se hace más significativa con el aumento de α . Esto se puede ver más claramente en la Fig. 4.7, que muestra la varianza de los multipolos de Legendre $\xi_{0,2,4}(s)$ a una escala de $s = 112,5 h^{-1}\text{Mpc}$, que coincide aproximadamente con la posición del pico de las BAO, recuperado a partir del estimador LS y la Ec. 4.9 en función del número total de puntos en los catálogos aleatorio y de los catálogos glass. Mientras que para $\alpha = 1$ la varianza de $\xi_0(s)$ es ~ 50 por ciento menor utilizando catálogos glass, es sólo ~ 25 por ciento la del LS para $\alpha = 8$. La desviación estándar de ambos estimadores tiene una dependencia de ley de potencia con respecto a N_r pero con diferentes pendientes. Mientras que la varianza disminuye a medida que $\sigma_{\xi_\ell} \propto N_r^{-0.53}$ para $\xi_{\text{LS}}(\mathbf{s})$, va como $\sigma_{\xi_\ell} \propto N_r^{-0.86}$ para $\xi_{\text{LS,G}}(\mathbf{s})$. Esto significa que una precisión deseada de los multipolos estimados se puede lograr utilizando catálogos glass que son significativamente más pequeños que las muestras aleatorias que se necesitarían con el enfoque LS estándar.

Estimadores adicionales de las funciones de correlación de dos y tres puntos

En esta sección damos versiones alternativas de los estimadores más comunes de funciones de correlación de dos y N puntos, adaptados al uso de catálogos glass en lugar de las distribuciones aleatorias estándar.

El estimador más sencillo de la función de dos puntos es el llamado estimador natural (Peebles y Hauser, 1974), que puede modificarse para utilizar dos catálogos glass como

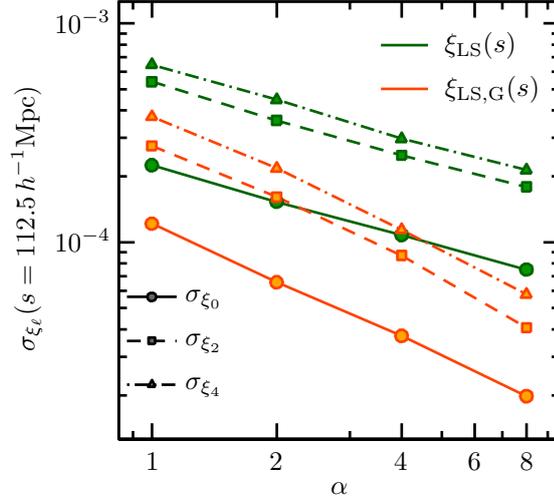


Figura 4.7: Desviaciones estándar σ_{ξ_ℓ} recuperadas a partir del estimador LS (líneas verdes) y la versión modificada de la Ec. 4.9 (líneas azules) en $s = 112.5 h^{-1} \text{Mpc}$ en función del número total de puntos en los catálogos aleatorio y glass. En ambos casos, la desviación estándar tiene una dependencia de ley de potencia del tamaño de los catálogos de referencia N_r , con $\sigma_{\xi_\ell} \propto N_r^{-0.53}$ para $\xi_{\text{LS}}(\mathbf{s})$ y como $\sigma_{\xi_\ell} \propto N_r^{-0.86}$ para $\xi_{\text{LS,G}}(\mathbf{s})$.

$$\xi_{\text{PH,G}}(\mathbf{s}) = \frac{DD(\mathbf{s})}{G_1 G_2(\mathbf{s})} - 1. \quad (4.13)$$

El estimador de Davis y Peebles (1983) puede implementarse con un único glass como

$$\xi_{\text{DP,G}}(\mathbf{s}) = \frac{DD(\mathbf{s})}{DG(\mathbf{s})} - 1. \quad (4.14)$$

Otro estimador comúnmente utilizado es el de Hamilton (1993b), que puede adaptarse al caso de dos catálogos glass como

$$\xi_{\text{H,G}}(\mathbf{s}) = \frac{DD(\mathbf{s})G_1 G_2(\mathbf{s})}{DG_1(\mathbf{s}) DG_2(\mathbf{s})} - 1. \quad (4.15)$$

La estimación de la función de correlación bipuntual tras la aplicación de la reconstrucción de Zeldovich requiere el uso de dos catálogos aleatorios, uno siguiendo la función de selección original, y un segundo, denotado por S , que se desplaza aplicando el mismo campo de desplazamiento que a los datos D . El estimador de la Ec. 4.7 se modifica entonces como (Padmanabhan et al., 2012)

$$\hat{\xi}_{\text{rec}}(\mathbf{s}) = \frac{DD(\mathbf{s}) - 2DS(\mathbf{s}) + SS(\mathbf{s})}{RR(\mathbf{s})}. \quad (4.16)$$

La adaptación de este estimador al uso de catálogos glass requiere cuatro muestras glass diferentes, dos de las cuales siguen los datos originales, $G_{1,2}$, y dos en las que se ha aplicado el campo de desplazamiento $G_{rmS1,2}$, lo que lleva a

$$\hat{\xi}_{\text{rec,G}}(\mathbf{s}) = \frac{DD(\mathbf{s}) - DG_{S1}(\mathbf{s}) - DG_{S2}(\mathbf{s}) + G_{S1}G_{S2}(\mathbf{s})}{G_1G_2(\mathbf{s})}. \quad (4.17)$$

Los estimadores de las funciones de correlación de orden superior también pueden adaptarse para utilizar catálogos glass. En la notación general de Szapudi y Szalay (1998), el estimador para la función de correlación de N puntos puede escribirse como

$$\xi_{N,G} = (D - G_1)(D - G_2) \cdots (D - G_N) / G_1G_2 \cdots G_N, \quad (4.18)$$

que requiere N catálogos glass independientes G_i . Para la función de correlación bipuntual, esta expresión corresponde al estimador LS modificado de la Ec. 4.9. Para la función de correlación de tres puntos, este estimador puede expresarse como

$$\begin{aligned} \xi_{3,G}(s_{12}, s_{23}, s_{13}) &= \frac{DDD(s_{12}, s_{23}, s_{13})}{G_1G_2G_3(s_{12}, s_{23}, s_{13})} \\ &\quad - \frac{\sum_i DDG_i(s_{12}, s_{23}, s_{13})}{G_1G_2G_3(s_{12}, s_{23}, s_{13})} \\ &\quad + \frac{\sum_{(i,j) \in \binom{\mathcal{I}}{2}} DG_iG_j(s_{12}, s_{23}, s_{13})}{G_1G_2G_3(s_{12}, s_{23}, s_{13})} + 1, \end{aligned} \quad (4.19)$$

donde $\mathcal{I} = \{1, 2, 3\}$, y $\binom{\mathcal{I}}{2}$ representa el conjunto de todas las posibles combinaciones de dos elementos de \mathcal{I} .

Los estimadores basados en catálogos glass pueden reducir significativamente la varianza de las mediciones de los estadísticos de alto orden con respecto a los resultados obtenidos con un único catálogo aleatorio con el mismo número total de puntos. El uso de catálogos glass más pequeños pueden entonces reducir significativamente el costo computacional total de la estimación de funciones de correlación de tres puntos sin afectar a la precisión de las mediciones.

4.5. Discusión

Hemos estudiado el impacto sobre la precisión de los estimadores al sustituir el catálogo aleatorio utilizado en la estimación de las funciones de correlación de dos puntos por distribuciones de puntos tipo glass. Mientras que el espectro de potencia de estas muestras se asemeja al de una distribución de Poisson en escalas menores que la separación media entre partículas, en escalas mayores sigue la forma mínima $P(k) \propto k^4$ y muestra una potencia significativamente menor que un catálogo aleatorio con el mismo número de puntos. A su vez, la varianza de los recuentos en esferas de radio R decae como $\sigma^2(R) \propto R^{-4}$, frente a $\sigma^2(R) \propto R^{-3}$ para el caso de Poisson, lo que reduce drásticamente el ruido en los recuentos de pares necesarios para estimar $\xi(\mathbf{s})$.

Hemos demostrado que se pueden generar distribuciones de partículas con las propiedades deseadas aplicando iterativamente la reconstrucción de Zeldovich a un conjunto de puntos inicialmente aleatorio. Esta tarea puede aprovechar los algoritmos rápidos que se han desarrollado en los últimos años en el contexto de la reconstrucción de las BAO (Burden et al., 2015). Aunque no hemos realizado un análisis detallado del número de iteraciones necesarias para obtener el catálogo glass óptimo para las mediciones de la función de correlación, hemos visto que, en las escalas consideradas aquí, la varianza de estas estimaciones converge después de tan sólo 5 iteraciones. El pequeño costo de cómputo adicional asociado a la construcción de dicha muestra se ve compensado por la mejora significativa en la precisión de las estimaciones de la función de correlación.

Hemos proporcionado una versión modificada del estimador LS, adaptada al uso de distribuciones de partículas tipo glass (Ec. 4.9). Este estimador hace uso de dos catálogos glass independientes para evitar problemas debidos a la correlación de los puntos dentro de una misma muestra para escalas que se acercan a la separación media entre partículas. El mismo estimador puede implementarse utilizando dos catálogos aleatorios diferentes (Ec. 4.10), ofreciendo un punto de referencia ideal para probar las ventajas de utilizar distribuciones de partículas tipo glass sobre las muestras de Poisson. Como estadístico de control, utilizamos las mediciones de los multipolos de Legendre $\xi_\ell(s)$ con $\ell = 0, 2, 4$ de muestras HOD que coinciden con las propiedades de agrupación de la muestra BOSS CMASS obtenida con diferentes realizaciones de catálogos aleatorios y glass.

Hemos encontrado que la varianza en gran escala del estimador de la Ec. 4.10 utilizando dos catálogos aleatorios separados es similar a la del estimador LS

estándar basado en un único conjunto con el mismo número total de puntos. Este estimador puede considerarse entonces como una forma sencilla de reducir el costo informático de la medición de $\xi_\ell(s)$ sin que ello afecte a la precisión de los resultados, de forma similar al método “split–random” de Keihänen et al. (2019).

Nuestros resultados muestran que el uso de las distribuciones tipo glass no añade ningún sesgo respecto a los resultados obtenidos utilizando muestras Poisson del mismo tamaño. En escalas más pequeñas que la separación media entre partículas de las muestras de referencia utilizando catálogos glass se obtiene un σ_{ξ_ℓ} similar al obtenido utilizando las distribuciones de Poisson. Sin embargo, en escalas mayores conducen a una reducción significativa de la varianza de $\xi_\ell(s)$ con respecto a los resultados del estimador LS estándar con el mismo número de puntos. Además, esta reducción se hace más grande con el aumento de $\alpha \equiv N_r/N_d$. El tamaño del catálogo glass necesario para lograr una precisión determinada en la función de correlación es significativamente menor que cuando se utilizan muestras aleatorias. Como el costo computacional de estos estimadores es proporcional al número total de pares, el menor tamaño del catálogo glass puede representar una reducción significativa del tiempo computacional total. Por ejemplo, extrapolando el comportamiento de ley de potencia de σ_{ξ_ℓ} mostrado en al Fig. 4.7 encontramos que la misma varianza conseguida por el estimador LS utilizando muestras aleatorias de $\alpha = 50$ como se hizo en los análisis finales del BOSS (Alam et al., 2017, Sánchez et al., 2017) se puede obtener utilizando catálogos glass con $\alpha = 5.4$. Una estimación más exigente basada en catálogos aleatorios con $\alpha = 100$ se puede igualar utilizando una muestra glass con $\alpha = 8.3$. El tiempo total de cálculo, t , de los estimadores de las Ecs. (4.7) y (4.9) está dominado por los términos RR y GG , y por tanto $t \propto \alpha^2$. Entonces, en los dos ejemplos anteriores, la reducción del tiempo de cálculo obtenida al utilizar catálogos glass en lugar de catálogos aleatorios asciende a $t_R/t_G \simeq 86$ y $t_R/t_G \simeq 145$, respectivamente, lo que representa una drástica reducción de los recursos computacionales necesarios para el análisis. Esta mejora puede ser especialmente beneficiosa para las mediciones de las BAO posteriores a la reconstrucción, que requieren recuentos de pares en catálogos aleatorios con y sin la aplicación del campo de desplazamiento.

Aunque nos hemos centrado en el estimador LS, la sección 4.4 contiene versiones de otros estimadores comúnmente utilizados de la $\xi(\mathbf{s})$ y de N -puntos adaptados al uso de catálogos glass. Los estimadores basados en muestras glass podrían resultar cada vez más útiles en el cálculo de estadísticas de alto orden, donde

el costo computacional de contar N -tuplas de puntos podría reducirse drásticamente sin comprometer la varianza. Aunque el tamaño del catálogo aleatorio no es el factor principal para determinar el costo computacional de las mediciones del espectro de potencia, la estadística del espacio de Fourier también podría beneficiarse de la menor varianza de los catálogos glass tanto en la estimación de $P(k)$ como en la función ventana del relevamiento. En los próximos años, relevamientos como DESI y Euclid proporcionarán muestras de decenas de millones de objetos en grandes volúmenes. El análisis de estos catálogos representará un reto para las técnicas de análisis tradicionales. El uso de catálogos glass podría ayudar a reducir drásticamente los requisitos computacionales de los análisis de agrupación en estos estudios y sus catálogos simulados asociados, manteniendo la alta precisión que exigen.

Capítulo 5

Nuevos estadísticos para el estudio de alineamientos

Parte de los objetivos de esta Tesis, como hemos establecido, es el esfuerzo por aportar nuevos métodos para estudios estadísticos de galaxias. En el capítulo anterior presentamos un tratamiento de catálogos aleatorios para reducir la varianza en el cálculo de funciones de correlación. A continuación, pasando de un escenario de estructura en gran escala a uno un poco más pequeño y astrofísico aunque igualmente estadístico, presentaremos dos métodos para recuperar señales de alineamiento en poblaciones de galaxias.

La motivación surge de las insuficiencias del estadístico usual para estos problemas: la mediana de la distribución de cosenos de los ángulos. Para una población con orientaciones aleatorias se espera una distribución uniforme de cosenos, y las características de una distribución uniforme las hace un estimador problemático. En particular, la desviación estándar de esta distribución es del orden de su media. Esto dificulta la obtención de una alta significación estadística de alguna señal encontrada de alineamiento, y determina que dependamos de simulaciones Monte Carlo para el cálculo de la misma.

En este capítulo, basado en el trabajo [Dávila-Kurbán et al. \(2022\)](#), presentamos dos métodos formales para analizar las alineaciones de una muestra de vectores con respecto a un centro. El primer método consiste en la definición de métricas simples a partir de las componentes radiales y tangenciales de los vectores, mientras que el segundo se basa en la parametrización de una función residual entre los datos obtenidos de la muestra y de una distribución isotrópica. En ambos casos, no asumimos ningún esquema de discretización (o *bineado*). En

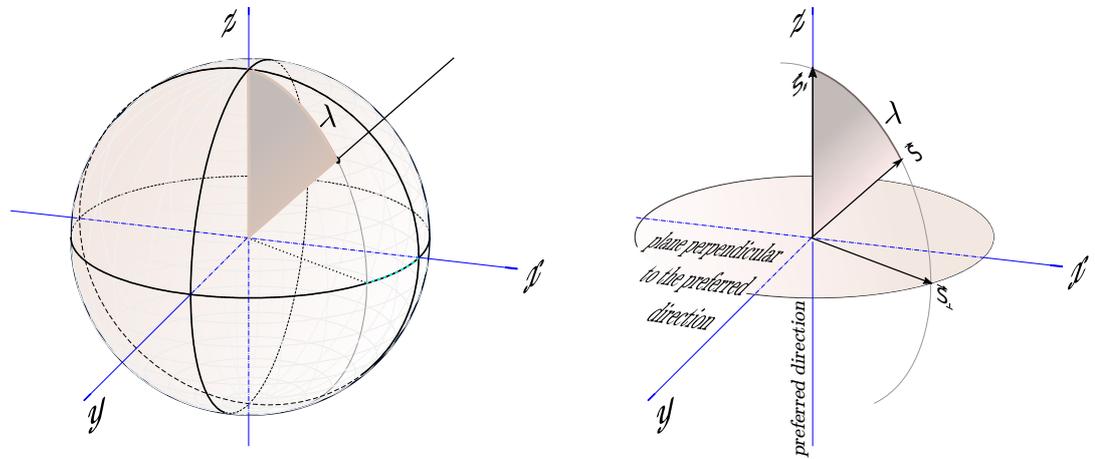


Figura 5.1: Sistema de coordenadas utilizado. El eje z es la dirección radial hacia el exterior del vacío. El ángulo θ se forma entre z y el vector \mathbf{S} , y toma valores en el rango $[0, \pi]$.

su lugar, utilizamos toda la información de los datos y aplicamos estimaciones robustas de las incertidumbres en las métricas de los alineamientos. Al derivar la distribución teórica de los parámetros que miden la señal de alineación, no sólo podemos determinar su significación estadística con precisión, sino que podemos hacerlo sin invertir tiempo y recursos computacionales en las simulaciones Monte Carlo que suelen ser necesarias para estimarlas.

5.1. Metodología y nuevos estadísticos

La simetría de estructuras como filamentos, halos, cúmulos o vacíos, tanto en su geometría como en su dinámica, permite considerar una dirección privilegiada para analizar la orientación de las galaxias. En el caso de sistemas con simetría esférica, ésta es la dirección radial. El objetivo aquí es desarrollar un formalismo estadístico para medir de forma robusta la distribución de la orientación de las galaxias y detectar posibles excesos con respecto a una distribución completamente aleatoria. Dado el problema de las orientaciones vectoriales con respecto a un punto central queremos definir un parámetro estadístico y obtener su distribución para conocer la significación de una prueba de hipótesis.

Relación de componentes vectoriales

Dada una dirección radial $\hat{\mathbf{z}}$ de norma unitaria (ver Fig. 5.1), las componentes perpendiculares y paralelas del vector \mathbf{S} pueden calcularse como

$$\mathbf{S}_{\parallel} = \mathbf{S} \cdot \hat{\mathbf{z}}, \quad \text{y} \quad \mathbf{S}_{\perp} = \mathbf{S} - \mathbf{S}_{\parallel}, \quad (5.1)$$

donde \mathbf{S}_{\perp} es la componente perpendicular a la dirección radial $\hat{\mathbf{z}}$, \mathbf{S}_{\parallel} es la componente paralela a la dirección radial y $\mathbf{S} = \mathbf{S}_{\perp} + \mathbf{S}_{\parallel}$.

El ángulo θ que forman la dirección radial y la dirección del vector \mathbf{S} se relaciona con las componentes:

$$S_{\perp} = |\mathbf{S}| \operatorname{sen}(\theta); \quad S_{\parallel} = |\mathbf{S}| \operatorname{cos}(\theta). \quad (5.2)$$

La distribución de este ángulo puede utilizarse para analizar las alineaciones, y dada su relación con los componentes, ésta también puede utilizarse para determinar las orientaciones. Para ello definimos:

$$\mathcal{B} = \frac{S_{\perp}}{S_{\parallel}} = \frac{S \operatorname{sen}(\theta)}{S \operatorname{cos}(\theta)} = \tan(\theta). \quad (5.3)$$

El parámetro \mathcal{B} es también una medida de la orientación del vector \mathbf{S} . Nótese que los rangos de estos dos parámetros son los siguientes:

$$0 \leq \theta \leq \pi; \quad -\infty \leq \mathcal{B} \leq \infty.$$

Usando la simetría del problema podemos definir parámetros considerando el ángulo agudo entre las direcciones $\hat{\mathbf{z}}$ y $\hat{\mathbf{S}}$, y la norma de la componente \mathbf{S}_{\parallel} ,

$$\beta = |\mathcal{B}| = \frac{S_{\perp}}{|S_{\parallel}|}, \quad \text{con} \quad \lambda = \min(\theta, \pi - \theta), \quad (5.4)$$

para los cuales:

$$0 \leq \lambda \leq \frac{\pi}{2}; \quad 0 \leq \beta \leq \infty.$$

Los vectores con $\beta > 1$ tienen preferencia de orientación en la dirección perpendicular, es decir $\frac{\pi}{4} < \lambda < \frac{\pi}{2}$, mientras que vectores con $\beta < 1$ se orientan preferentemente en la dirección radial ($\hat{\mathbf{z}}$ en la Fig. 5.1), con $0 < \lambda < \frac{\pi}{4}$.

Para cuantificar la dirección de \mathbf{S} se podrían considerar las siguientes estadísticas:

- el ángulo θ
- el ángulo agudo λ
- el cociente entre los componentes perpendiculares y paralelos, \mathcal{B} (con θ)
- el cociente entre los componentes perpendiculares y paralelos, β (con λ)

En las siguientes subsecciones exploraremos el uso de los ángulos o los cocientes. Es de suma importancia establecer las distribuciones de estos parámetros para el caso en el que no hay ninguna señal de alineación. De este modo, determinamos la amplitud de las fluctuaciones estadísticas y establecemos una medida de la señal en una muestra de datos calculando su significación estadística. Para ello, definimos la hipótesis nula

H_0 : la distribución de los vectores es aleatoria con simetría esférica

es decir, no hay ninguna señal de alineación. Esta hipótesis también puede utilizarse para generar muestras de control con procedimientos de Monte Carlo si es necesario.

Nótese que las regiones de β mayores o menores que 1 son diferentes, por lo que se espera que para una distribución aleatoria haya más vectores “perpendiculares” que vectores “paralelos” (véase la Fig. 5.2, panel superior).

5.1.1. Cociente de los componentes vectoriales, β

En esta sección derivamos la distribución del estadístico de prueba β , definido como la relación entre las componentes de los vectores perpendicular y paralelo (Ec. 5.4).

La distribución de β puede deducirse del teorema del cambio de variables aleatorias, que, en su forma general, puede enunciarse así (Gillespie, 1983):

Sea $\{X_i\}_{i=1}^n$ una variable aleatoria (V.A.) con $f_{\mathbf{X}}(\mathbf{x})$ conocida, y sea $\mathbf{Y} = \varphi(\mathbf{x})$ una V.A. donde $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_m)$, siendo $\varphi_k : \mathbb{R}^n \rightarrow \mathbb{R}$ funciones reales. La función de probabilidad conjunta $f_{\mathbf{Y}}(\mathbf{y})$ viene dada por:

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_{-\infty}^{\infty} d\mathbf{x} f_{\mathbf{X}}(\mathbf{x}) \prod_{i=1}^m \delta(y_i - \varphi_i(\mathbf{x})), \quad (5.5)$$

donde δ es la función Delta de Dirac. Para el caso particular de una variable unidimensional, $X : \Omega \rightarrow \mathbb{R}$ y $Y : \Omega \rightarrow \mathbb{R}$, con $Y = \varphi(X)$,

$$f_Y(y) = \int_{-\infty}^{\infty} dx f_X(x) \delta(y - \varphi(x)). \quad (5.6)$$

Entonces, podemos utilizar este teorema para encontrar la distribución de β partiendo de $F_X(x) = U(0, 1)$ con la transformación:

$$\beta = \tan(\arccos(x)), \quad (5.7)$$

o bien a partir de $f_\Lambda(\lambda) = \text{sen}(\lambda)$ con la transformación:

$$\beta = \tan(\lambda), \quad 0 < \lambda < \pi/2. \quad (5.8)$$

Usando esta última, tenemos:

$$\begin{aligned} f_B(\beta) &= \int_{-\infty}^{\infty} d\lambda f_\Lambda(\lambda) \delta(\beta - \tan(\lambda)) \\ &= \int_0^{\pi/2} d\lambda \text{sen}(\lambda) \delta(\beta - \tan(\lambda)). \end{aligned} \quad (5.9)$$

Para resolver esta integral, hacemos el cambio de variables:

$$z = \tan(\lambda) \implies \lambda = \arctan(z), \quad d\lambda = \frac{dz}{1+z^2}.$$

Por lo tanto,

$$f_B(\beta) = \int_0^{\infty} dz \frac{\text{sen}(\arctan(z))}{1+z^2} \delta(\beta - z) = \frac{\text{sen}(\arctan(\beta))}{1+\beta^2}. \quad (5.10)$$

Esta expresión se puede simplificar utilizando las propiedades de las funciones trigonométricas. En efecto, si $\beta = \tan(z)$ para un número z , entonces:

$$\begin{aligned} \beta^{-2} + 1 &= \frac{1}{\tan(z)^2} + 1 = \frac{\cos(z)^2}{\text{sen}(z)^2} + 1 = \frac{\text{sen}(z)^2 + \cos(z)^2}{\text{sen}(z)^2} \\ &= \frac{1}{\text{sen}(z)^2} \end{aligned}$$

$$\begin{aligned}
\Rightarrow \frac{1}{\text{sen}(z)} &= \sqrt{\beta^{-2} + 1} = \sqrt{\frac{1 + \beta^2}{\beta^2}} = \frac{\sqrt{1 + \beta^2}}{\beta} \\
\Rightarrow \text{sen}(z) &= \frac{\beta}{\sqrt{1 + \beta^2}} \\
\Rightarrow \text{sen}(\arctan(\beta)) &= \frac{\beta}{\sqrt{1 + \beta^2}}.
\end{aligned}$$

Reemplazando en la Ec. 5.10 tenemos,

$$\begin{aligned}
f_B(\beta) &= \frac{\text{sen}(\arctan(\beta))}{1 + \beta^2} \\
&= \frac{\beta}{\sqrt{1 + \beta^2}} \frac{1}{1 + \beta^2} \\
&= \beta(1 + \beta^2)^{-3/2}.
\end{aligned} \tag{5.11}$$

La función de probabilidad es por lo tanto:

$$\begin{aligned}
F_B(\beta) &= \int_0^\beta f_B(b) db \\
&= \int_0^\beta b(1 + b^2)^{-3/2} db \\
&= -\frac{1}{\sqrt{1 + b^2}} \Big|_0^\beta \\
&= 1 - \frac{1}{\sqrt{1 + \beta^2}}.
\end{aligned} \tag{5.12}$$

La Fig. 5.2, panel superior, muestra la distribución teórica de β , con el muestreo Monte Carlo de la variable aleatoria mostrada con el histograma.

Conociendo la distribución f_B se pueden realizar análisis de las orientaciones de los vectores con respecto a una dirección determinada. En general, no es útil medir un único valor de la V.A. β , dado que está sujeto a fluctuaciones aleatorias. Por tanto, calculamos los valores del estimador β en una muestra de observaciones. Es decir, analizamos una muestra aleatoria (M.A.) de valores para determinar si difiere de los resultados esperados para una distribución aleatoria (una M.A. bajo

la hipótesis nula) de vectores. Para formalizar estos análisis necesitamos establecer algunas propiedades básicas de la distribución f_B .

El primer momento de la distribución, si existe, es:

$$\begin{aligned} E[B] &= \int_0^\infty t f_B(t) dt \\ &= \int_0^1 t f_B(t) dt + \int_1^\infty t f_B(t) dt, \end{aligned} \quad (5.13)$$

donde

$$\int_1^\infty t f_B(t) dt = \int_1^\infty t \frac{t}{(1+t^2)^{3/2}} dt.$$

Teniendo en cuenta que para un número real $x > 1$ tenemos $x^n > x$, y $x > \sqrt{x}$, por tanto $x^{3/2} = x\sqrt{x} < x$. Entonces, para $\beta > 1$, $1 + \beta^2 > 2 > 1$ y sabiendo que

$$\frac{1}{(1+t^2)^{3/2}} > \frac{1}{(1+t^2)},$$

luego podemos acotar la integral:

$$\begin{aligned} \int_1^\infty t \frac{t}{(1+t^2)^{3/2}} dt &> \int_1^\infty \frac{t^2}{(1+t^2)} dt \\ &> \int_1^\infty \frac{t}{(1+t^2)} dt \\ &> \lim_{M \rightarrow \infty} \frac{1}{2} \ln(1+t^2) \Big|_1^M = \infty. \end{aligned}$$

Vemos, entonces, que el valor de expectación $E[B]$ es indefinido. En efecto, ningún momento de esta distribución está definido. De hecho, teniendo en cuenta que:

$$\begin{aligned} E[B^n] &= \int_0^\infty t^n f_B(t) dt \\ &= \int_0^1 t^n f_B(t) dt + \int_1^\infty t^n f_B(t) dt, \end{aligned} \quad (5.14)$$

y que:

$$\beta > 1 \implies 1 + \beta^1 > 1 \implies \frac{\beta^n}{1 + \beta^2} > \frac{\beta}{1 + \beta^2}$$

para $n \geq 1$. Luego,

$$\int_1^\infty t^n f_B(t) dt > \int_1^\infty t f_B(t) dt > \infty.$$

La distribución $f_B(\beta)$ es una distribución patológica donde los momentos están indefinidos. Esta limitación impide utilizar procedimientos de Monte Carlo para estimar la distribución de $\bar{\beta}$ porque no es posible asegurar que los valores medios de β sigan una distribución estable. Para trabajar con esta distribución, podríamos idear una distribución truncada entre valores arbitrarios L_1 y L_2 , con la condición de que $L_1 \sim 0$ y L_2 sea mucho mayor que la región de interés del parámetro β , que es la región alrededor de $\beta = 1$.

Por ejemplo, si elegimos

$$L_1 = 10^{-3}; \quad L_2 = 10^3$$

resulta que, definiendo el factor de corrección κ :

$$\kappa = \int_{L_1}^{L_2} f_B(t) dt = \frac{1}{\sqrt{1+L_1^2}} - \frac{1}{\sqrt{1+L_2^2}} = \frac{1}{\sqrt{1+10^{-6}}} - \frac{1}{\sqrt{1+10^6}}$$

podemos definir una función de distribución aproximada para B , definida como:

$$f_{\tilde{B}}(\beta) = \begin{cases} \frac{1}{\kappa} f_B(\beta) & \beta \in [L_1, L_2] \\ 0 & \text{c.c.,} \end{cases}$$

En efecto, la media de esta función está definida y su expresión es la siguiente:

$$E[\tilde{B}] = \frac{1}{\kappa} \left[\frac{L_1}{\sqrt{L_1^2+1}} - \frac{L_1^2 \operatorname{asinh}(L_1)}{L_1^2+1} + \frac{L_2^2 \operatorname{asinh}(L_2)}{L_2^2+1} - \frac{L_2}{\sqrt{L_2^2+1}} + \frac{\operatorname{asinh}(L_2)}{L_2^2+1} - \frac{\operatorname{asinh}(L_1)}{L_1^2+1} \right].$$

Si tomamos $L_1 = 1/L_2$, tenemos:

$$E[\tilde{B}] = \frac{1}{\kappa} \left[\frac{1/L_2}{\sqrt{L_2^{-2} + 1}} - \frac{L_2^{-2} \operatorname{asinh}(1/L_2)}{L_2^{-2} + 1} + \frac{L_2^2 \operatorname{asinh}(L_2)}{L_2^2 + 1} - \frac{L_2}{\sqrt{L_2^2 + 1}} + \frac{\operatorname{asinh}(L_2)}{L_2^2 + 1} - \frac{\operatorname{asinh}(1/L_2)}{L_2^{-2} + 1} \right]$$

Sin embargo, el resultado depende en gran medida del valor de L_2 y, en menor medida, del valor de L_1 . Sea

$$A(1/L_2, L_2) = \int_{1/L_2}^{L_2} t f_B(t) dt,$$

es sencillo ver que $A(1/L_2, L_2)$ depende de L_2 , donde tomamos diferentes valores para L_2 y $L_1 = 1/L_2$.

Con esto demostramos que no es posible obtener una distribución para $\bar{\beta}$. Además, no sólo no es posible resolverla analíticamente, sino que tampoco es posible hacer una estimación bootstrap formal del error. Sin embargo, como mostramos en la siguiente sección, β puede utilizarse para definir un nuevo parámetro con mejores propiedades estadísticas.

5.1.2. Cociente entre vectores perpendiculares y paralelos, η

Para encontrar un estimador robusto consideramos la fracción de valores de β que son mayores que algún valor crítico. Dado que cuando no hay preferencia en la orientación, las componentes perpendicular y paralela son iguales, podemos plantear que dicho valor crítico sea $\beta = 1$. Por tanto, definimos un parámetro que representa la fracción en exceso de vectores alineados en una dada dirección, por ejemplo:

$$\hat{\eta} = \frac{N(\beta > 1)}{N(\beta < 1)}, \quad (5.15)$$

donde N es el número de observaciones de una muestra que cumple las condiciones indicadas entre paréntesis. Bajo H_0 , en base a la función de densidad de probabilidad, se espera que

$$\eta_0 = \frac{P(\beta > 1)}{P(\beta < 1)} \quad (5.16)$$

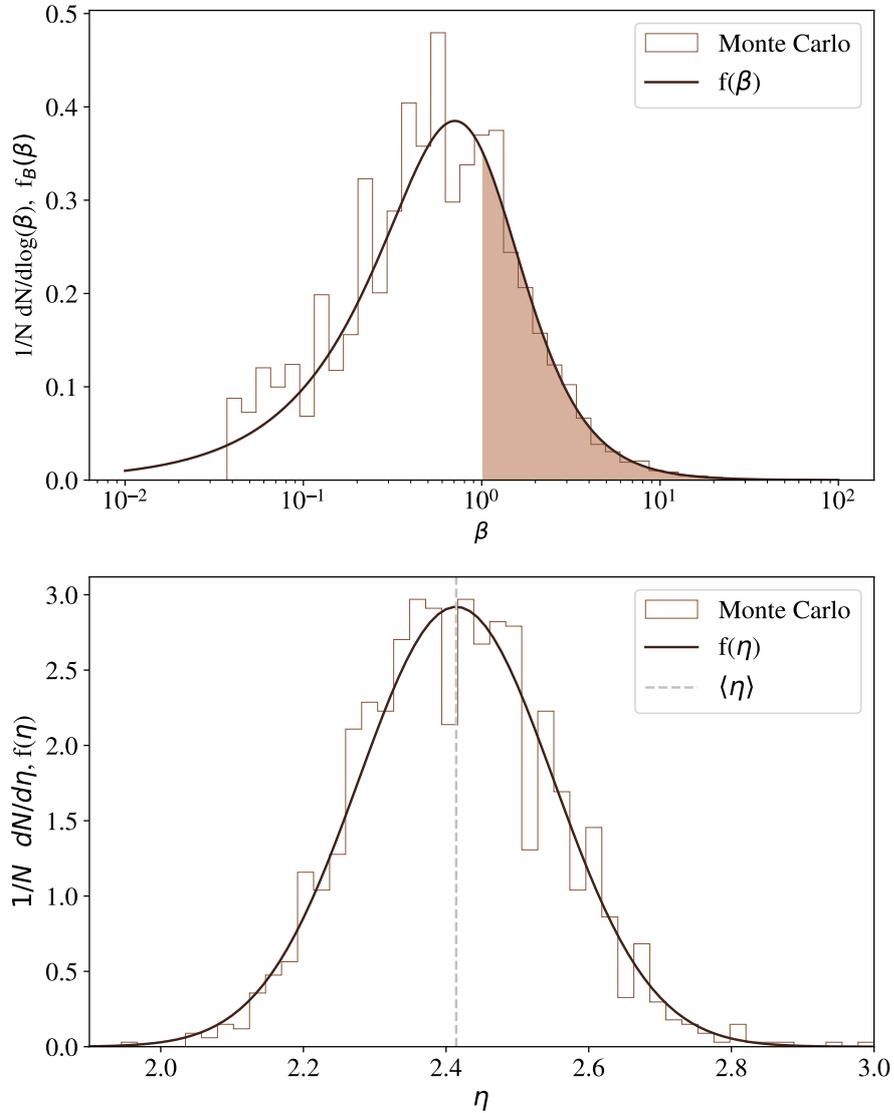


Figura 5.2: (superior:) Distribución de β obtenida a partir de la derivación teórica (línea sólida) y de las simulaciones de Monte Carlo (histograma), y definición de η (fracción de realizaciones con $\beta > 1$). (abajo:) Histograma de las variables η ordenadas con el método de Montecarlo (a partir de muestras de β) y con la aproximación de la distribución teórica. El valor teórico medio ($1/(\sqrt{2} - 1)$) se muestra en la línea discontinua y el histograma corresponde a una realización de Monte Carlo de los valores de η .

Para calcular el valor de η_0 , tenemos en cuenta que utilizando la función de probabilidad, F_B (ver Ec. 5.12):

$$P(\beta > 1) = 1 - F_B(1) = 1 - \left(1 - \frac{1}{\sqrt{1 + \beta^2}}\right) \Big|_{\beta=1} = \frac{1}{\sqrt{2}}$$

$$P(\beta < 1) = F_B(1) = 1 - \frac{1}{\sqrt{1 + \beta^2}} \Big|_{\beta=1} = 1 - \frac{1}{\sqrt{2}}$$

i.e.:

$$\begin{aligned} \eta_0 &= \frac{P(\beta > 1)}{P(\beta < 1)} = \frac{\int_0^1 f_B(t) dt}{\int_1^\infty f_B(t) dt} = \frac{\frac{1}{\sqrt{2}}}{1 - \frac{1}{\sqrt{2}}} = \frac{1}{\sqrt{2} - 1} \\ &\cong 2.4142. \end{aligned} \tag{5.17}$$

Proponemos que $\hat{\eta}$ sea un estimador insesgado de η_0 , es decir tenemos que comprobar si $E(\hat{\eta}) = \eta$. Sea una muestra aleatoria de N valores de β , definimos:

$$n = N(\beta > 1)$$

Dado que la probabilidad de obtener un valor de $\beta > 1$ es $P(\beta > 1) = 1/\sqrt{(2)}$, la variable n tiene una distribución Binomial,

$$f_n(n) = \text{Bin}(p, N) = \binom{N}{n} p^n (1-p)^{N-n}$$

con $p = 1/\sqrt{2} \approx 0.707$. Por lo tanto,

$$\eta = \frac{n}{N - n},$$

y la distribución de η entonces se puede obtener a partir de la distribución de n , teniendo en cuenta que:

$$P_\eta \left(\eta = \frac{k}{N - k} \right) = P_n(n = k),$$

donde $k = \frac{\eta N}{\eta - 1}$.

Es equivalente entonces, aunque mucho más eficiente, generar con este método variables aleatorias de la distribución de η que con el método de Monte Carlo. La

comparación entre las dos muestras aleatorias puede verse en el panel inferior de la Fig. 5.2.

Entonces, teniendo en cuenta que el valor de expectación de la variable $n \sim Bin(M, p)$ es Mp , tenemos que calcular el valor de expectación del cociente. Este problema no suele estar bien definido, pero se puede resolver de forma aproximada. A continuación verificamos que η_0 es efectivamente el valor de expectación de η y derivamos una expresión analítica para su varianza.

Valor de expectación y varianza de η

Sean X, Y V.A. definidas como $X = n, Y = N - n$. Si $q = 1 - p$, los valores de expectación de estas variables son:

$$\mu_X = Np; \quad \mu_Y = N - Np = N - \mu_X$$

y las varianzas:

$$\sigma_X^2 = \sigma_Y^2 = NP(1 - p) = Npq$$

con $q = 1 - p$.

Queremos calcular el valor de expectación de la relación $Q = X/Y$. De la definición del valor de expectación se desprende que no es posible derivar una expresión simple para Q (Duris et al., 2018, Koopman, 1984, R.M. y Bonett, 2008). Otra razón que impide resolver analíticamente la distribución de Q es que el denominador puede ser cero. Una forma de resolver este problema es reescribir la función de forma que se evite tener una singularidad. Es posible realizar dicha aproximación a partir del desarrollo de la serie de Taylor de $\bar{Q}(X, Y) = X/Y$ alrededor de $(X, Y) = (\mu_X, \mu_Y)$, es decir, $\bar{Q} = Q + R$:

$$\begin{aligned} \bar{Q}(X, Y) &= \bar{Q}(\mu_X, \mu_Y) + \frac{\partial}{\partial X} \bar{Q}(\mu_X, \mu_Y)(X - \mu_X) \\ &\quad + \frac{\partial}{\partial Y} \bar{Q}(\mu_X, \mu_Y)(Y - \mu_Y) + R, \end{aligned}$$

donde R es el error de orden 2 dado por el teorema de Taylor. Entonces, podemos estimar el valor de expectación de \bar{Q} , $E[\bar{Q}] \approx E[Q]$, donde

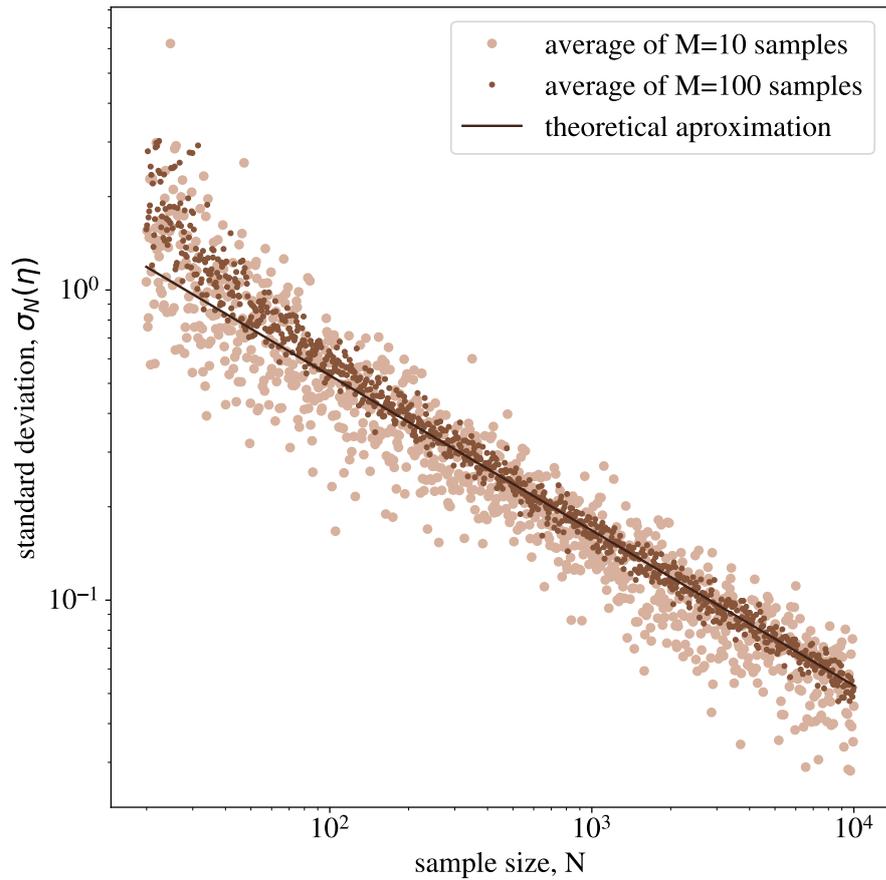


Figura 5.3: Variación de las estimaciones de Monte Carlo de la varianza de M muestras de valores de η , calculadas a partir de N muestras de valores de β (N , en el eje X, simulado), para $M=10$ (puntos grandes) y $M=100$ (puntos pequeños). Se demuestra que la varianza teórica es una estimación adecuada para muestras de tamaño $M > 100$.

$$\begin{aligned}
E[Q] &= E\left[Q(\mu_X, \mu_Y) + \frac{\partial Q}{\partial X}(\mu_X, \mu_Y)(X - \mu_X) + \frac{\partial Q}{\partial Y}(\mu_X, \mu_Y)(Y - \mu_Y)\right] \\
&= Q(\mu_X, \mu_Y) + \frac{\partial Q}{\partial X}(\mu_X, \mu_Y)E[(X - \mu_X)] + \frac{\partial Q}{\partial Y}(\mu_X, \mu_Y)E[(Y - \mu_Y)] \\
&= Q(\mu_X, \mu_Y).
\end{aligned}$$

Entonces, en una aproximación de primer orden y considerando $Q = \eta$, $X = n$, $Y = N - n$:

$$E(\hat{\eta}) \approx \frac{Np}{N - Np},$$

y

$$\frac{Np}{N - Np} = \frac{p}{1 - p} = \frac{\frac{1}{\sqrt{2}}}{1 - \frac{1}{\sqrt{2}}} = \frac{P(\beta > 1)}{P(\beta < 1)} = \eta_0,$$

por lo que

$$E(\hat{\eta}) \approx \eta_0.$$

El momento de segundo orden de $\hat{\eta}$, se obtiene calculando la varianza de la aproximación de primer orden de $Q(X, Y)$:

$$\begin{aligned}
Var(Q) &= Var \left[Q(\mu_X, \mu_Y) + \frac{\partial Q}{\partial X}(\mu_X, \mu_Y)(X - \mu_X) + \frac{\partial Q}{\partial Y}(\mu_X, \mu_Y)(Y - \mu_Y) \right] \\
&= \left(\frac{\partial Q}{\partial X}(\mu_X, \mu_Y) \right)^2 \sigma_X^2 + \left(\frac{\partial Q}{\partial Y}(\mu_X, \mu_Y) \right)^2 \sigma_Y^2 + \\
&\quad + \left(2 \frac{\partial Q}{\partial X}(\mu_X, \mu_Y) \frac{\partial Q}{\partial Y}(\mu_X, \mu_Y) \right) Cov[X, Y] \\
&= \left[\left(\frac{\partial Q}{\partial X}(\mu_X, \mu_Y) \right)^2 + \left(\frac{\partial Q}{\partial Y}(\mu_X, \mu_Y) \right)^2 \right] \sigma_X^2 + \\
&\quad + \left(2 \frac{\partial Q}{\partial X}(\mu_X, \mu_Y) \frac{\partial Q}{\partial Y}(\mu_X, \mu_Y) \right) Cov[X, Y].
\end{aligned}$$

Para evaluar las derivadas, tenemos que:

$$\begin{aligned}
\frac{\partial \eta}{\partial X}(\mu_X, \mu_Y) &= \frac{\partial(X/Y)}{\partial X}(\mu_X, \mu_Y) = \frac{1}{Y} \Big|_{(X,Y)=(\mu_X,\mu_Y)} \\
&= \frac{1}{\mu_Y} = \frac{1}{N(1-p)} = \\
&= \frac{1}{Nq},
\end{aligned} \tag{5.18}$$

y

$$\begin{aligned}
\frac{\partial \eta}{\partial Y}(\mu_X, \mu_Y) &= \frac{\partial X/Y}{\partial Y}(\mu_X, \mu_Y) = -\frac{X}{Y^2} \Big|_{(X,Y)=(\mu_X,\mu_Y)} \\
&= -\frac{\mu_X}{\mu_Y^2} = -\frac{Np}{N^2(1-p)^2} \\
&= -\frac{p}{Nq^2}.
\end{aligned} \tag{5.19}$$

La covarianza entre X e Y, está dada por:

$$\begin{aligned}
\text{Cov}(X, Y) &= \frac{1}{N} \sum_{i=1}^N (x_i - E(X))(y_i - E(Y)) \\
&= \frac{1}{N} \sum_{i=1}^N (x_i - \mu_X) \left((N - x_i) - (N - \mu_X) \right) \\
&= -\frac{1}{N} \sum_{i=1}^N (x_i - \mu_X)^2 \tag{5.20}
\end{aligned}$$

$$= -\sigma_X^2. \tag{5.21}$$

Luego, usando las expresiones 5.18, 5.19 y 5.21

$$\begin{aligned}
\text{Var}(\eta) &= \left[\left(\frac{1}{Nq} \right)^2 + \left(\frac{p}{Nq^2} \right)^2 \right] Npq + 2 \frac{1}{Nq} \frac{p}{Nq^2} Npq \\
&= \frac{1}{N} \left(\left(\frac{1}{q^2} + \frac{p^2}{q^4} \right) pq + 2 \frac{p^2}{q^2} \right) \\
&= \frac{1}{N} \left[\frac{p}{q} + 2 \left(\frac{p}{q} \right)^2 + \left(\frac{p}{q} \right)^3 \right]. \tag{5.22}
\end{aligned}$$

Evaluando en $p = 1/\sqrt{2}$, tenemos que

$$\frac{p}{q} = \frac{1/\sqrt{2}}{1 - 1/\sqrt{2}} = \sqrt{2} + 1.$$

Luego tenemos, finalmente,

$$\text{Var}(\eta) \simeq \frac{28.14214}{N}. \tag{5.23}$$

Se puede observar en la Fig. 5.3 que la varianza teórica de la expresión 5.23 y los valores de la varianza calculados a partir de 10 o 100 muestras de valores de η , obtenidas en muestras simuladas equivalentes de β con el método de la distribución binomial. Utilizar el valor teórico es equivalente a utilizar las simulaciones de Montecarlo cuando $N \gtrsim 100$, con la ventaja de no necesitar comparativamente ningún tiempo de cálculo.

5.1.3. Una prueba para los coeficientes OLS de la distribución del coseno

El segundo método que presentamos se basa en analizar la distribución de $\cos(\lambda)$ para determinar si es distinguible de la distribución esperada para una muestra de orientaciones aleatorias de vectores \mathbf{S} . Como se ha comentado, dicha distribución es uniforme bajo H_0 . Trabajando con muestras de tamaño limitado, las fluctuaciones estadísticas pueden generar diferencias entre los dos conjuntos de datos, incluso cuando surjan de la misma distribución. Por lo tanto, queremos comparar las dos distribuciones y establecer si su diferencia es suficiente para descartar H_0 .

La comparación entre dos distribuciones observadas se realiza de forma más robusta a partir de la distribución acumulativa empírica. Si se tiene una muestra aleatoria de una variable X , $\{x_i\}_{i=1}^n$, donde los valores están ordenados, la función de distribución acumulativa empírica (ECDF, por su nombre en inglés *empirical cumulative distribution function*) es:

$$F_e[x] = \frac{|\{X/X < x\}|}{|\{X\}|} \quad (5.24)$$

Esta función se utiliza en la prueba de Kolmogorov-Smirnov (Marsaglia et al., 2003), donde el estadístico D se define como la máxima diferencia entre las dos distribuciones acumulativas. Siguiendo esta idea, para describir la diferencia entre una distribución observada y una distribución de control $F_c(x)$, consideramos:

$$\Delta(x) = F_e[X](x) - F_c(x). \quad (5.25)$$

En el caso de la distribución del parámetro λ , sabemos que bajo H_0 tiene que ser uniforme entre 0 y $\pi/2$, lo que da como resultado $F_c(x) = 2x/\pi$. Entonces,

$$\Delta(x) = F_e[X](x) - \frac{2x}{\pi}. \quad (5.26)$$

Esta función, por definición, comienza y termina en cero, es decir,

$$\Delta(x) = 0 \text{ para } x = 0; x = \pi/2.$$

Para representar la función $\Delta(x)$ se puede utilizar una base de funciones ortogonales. Si $f(x)$ es una función continua en el intervalo $[0, \pi/2]$, entonces se puede

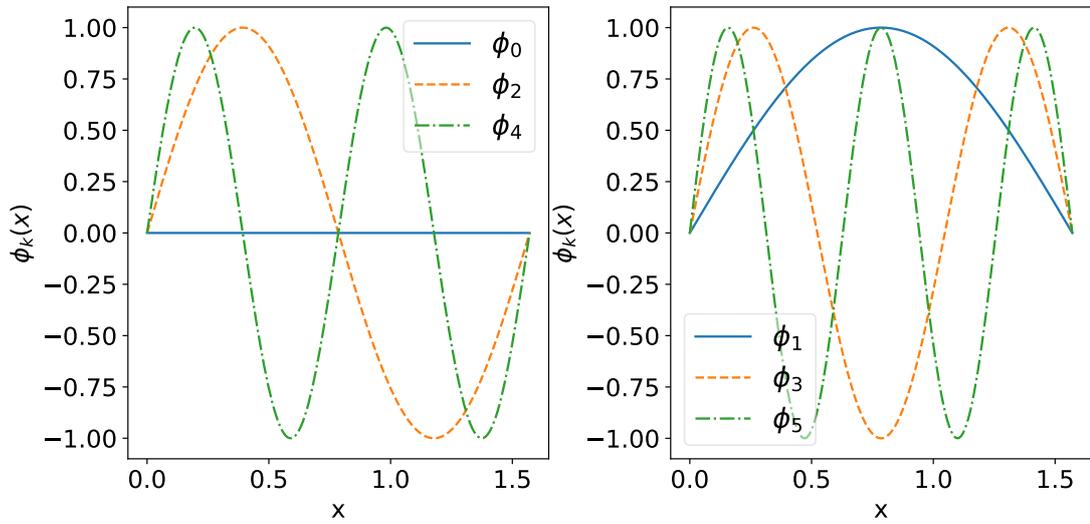


Figura 5.4: Primeros elementos pares e impares de la base de funciones $\phi(x)$.

escribir como

$$\Delta(x) = \sum_{k=0}^{\infty} a_k \phi_k(x), \quad (5.27)$$

donde

$$\phi_k(x) = \text{sen}(2kx).$$

Los primeros elementos de esta base se pueden ver en la Fig. 5.4.

Para definir la ortogonalidad de la base usamos el producto interno definido como:

$$\langle f, g \rangle = \int_L f(t)g(t)dt.$$

En este caso $L = [0, \pi/2]$, y

$$\begin{aligned} \langle \phi_m, \phi_n \rangle &= \int_0^{\pi/2} \phi_m(x)\phi_n(x)dx \\ &= \int_0^{\pi/2} \text{sen}(2mx) \text{sen}(2nx)dx, \end{aligned}$$

que se puede resolver teniendo en cuenta las relaciones

$$\begin{aligned}\cos(a + b) &= \cos(a) \cos(b) - \operatorname{sen}(a) \operatorname{sen}(b) \\ \implies \operatorname{sen}(a) \operatorname{sen}(b) &= \frac{1}{2} [\cos(a - b) - \cos(a + b)].\end{aligned}$$

Luego,

$$\begin{aligned}\langle \phi_m, \phi_n \rangle &= \int_0^{\pi/2} \operatorname{sen}(2mx) \operatorname{sen}(2nx) dx \\ &= \frac{1}{2} \int_0^{\pi/2} \cos(2mx - 2nx) - \cos(2mx + 2nx) dx \\ &= \frac{1}{2} \int_0^{\pi/2} \cos(2x(m - n)) dx - \frac{1}{2} \int_0^{\pi/2} \cos(2x(m + n)) dx.\end{aligned}$$

Teniendo en cuenta que $m - n$ es par si y solo si $m + n$ es par, resulta que para $m - n \neq 0$:

$$\langle \phi_m, \phi_n \rangle = 0,$$

mientras que

$$\langle \phi_n, \phi_n \rangle = \frac{1}{2} \int_0^{\pi/2} \cos(0) dx - \frac{1}{2} \int_0^{\pi/2} \cos(4xn) dx = \frac{\pi}{4}.$$

Por lo tanto, definiendo

$$\phi_k(x) = \langle \phi_n, \phi_n \rangle \operatorname{sen}(2kx),$$

tenemos un sistema ortonormal. Y podemos finalmente escribir:

$$\phi_k(x) = \frac{4}{\pi} \operatorname{sen}(2kx). \quad (5.28)$$

No obstante, en el caso de los conjuntos de datos, no existe una función continua, sino un muestreo discreto. El análisis de Fourier permite una expansión en términos de una suma finita de senos en el caso de un conjunto de puntos igualmente distanciados. Si los puntos no están igualmente distanciados debemos recurrir a estrategias alternativas. Se puede hacer una discretización, pero esto implica no hacer un uso óptimo de la información disponible. Otra opción es

ajustar un modelo a los puntos observados. Por ejemplo, podríamos tomar como modelo:

$$\Phi(x) = \sum_{k=0}^M a_k \phi_k(x), \quad (5.29)$$

que es un modelo lineal en los parámetros que queremos calcular: los coeficientes a_k .

Aunque este método es suficiente para representar una distribución, no es útil como estadística para caracterizar los parámetros de la distribución. Sin embargo, es posible aprovechar las condiciones de simetría para caracterizar la distribución.

Por ejemplo, si asumimos simetría, es decir, si tomamos como observables sólo los cosenos de los ángulos *agudos* entre la población de vectores y la dirección preferencial, esperamos que los parámetros pares sean cero. La Ec. 5.29 está modelando la diferencia entre una distribución acumulativa de la distribución del coseno calculada a partir de los datos y una función de control, por lo que se espera que al estudiar los efectos de los alineamientos preferenciales en una población de vectores esta función tome valores mayoritariamente positivos o negativos, indicando un alineamiento neto perpendicular o paralelo a la dirección preferida, respectivamente. Por lo tanto, en el caso de la simetría en la que nos interesa el ángulo agudo λ , se observa en el panel derecho de la Fig. 5.4 que el primer término, representado en una curva azul, será el término dominante en la Ec. 5.29 y es proporcional al parámetro a_1 ; el área bajo la curva, ya sea positiva o negativa, representará un alineamiento neto en la dirección perpendicular o paralela, respectivamente.

Los coeficientes a_k pueden obtenerse por mínimos cuadrados ordinarios a partir de los datos. Dada la Ec. 5.29 y los residuos de la ECDF de los datos $y_i = i/n - i$ de tamaño n , pretendemos minimizar:

$$\chi^2 = \sum_{i=1}^n \left(\frac{y_i}{\sigma_i} - \frac{1}{\sigma_i} \Phi(x) \right)^2 = |\mathbf{A}a_k - \mathbf{B}|, \quad (5.30)$$

donde \mathbf{A} y \mathbf{B} están en notación matricial y x son los datos que queremos ajustar, es decir, los cosenos calculados de la muestra. Así que minimizando esta expresión se deduce que $\mathbf{A}a_k$ es igual a \mathbf{B} (Hastie et al., 2001, Capítulo 3), por lo que:

$$a_k = \frac{(\mathbf{A}^T \mathbf{B})_k}{(\mathbf{A}^T \mathbf{A})_k}. \quad (5.31)$$

La base de las funciones es el conjunto de todas las funciones armónicas $\phi_k(x) = \text{sen}(k\pi x)$, que truncamos en $k=4$, y suponiendo $\sigma_i = 1$, tenemos:

$$(\mathbf{A}^T \mathbf{A})_k = \sum_{i=1}^n \phi_k^2(x_i). \quad (5.32)$$

Por otro lado, tenemos $(\mathbf{A}^T \mathbf{B})_k = \sum_{i=1}^n \phi_k(x_i) y_i$, por lo que, tras sustituir tenemos:

$$a_k = \frac{\sum_{i=1}^n \phi_k(x_i) y_i}{\sum_{i=1}^n \phi_k^2(x_i)}. \quad (5.33)$$

En nuestro caso de estudio tenemos $x = \cos(\lambda)$, y suponiendo datos ordenados: $y_i = i/n - i$, finalmente llegamos a una expresión analítica para los coeficientes obtenidos por el ajuste de mínimos cuadrados ordinarios (OLS, por sus siglas en inglés):

$$a_k = \frac{\sum_{i=1}^n i \text{sen}[k\pi \cos(\lambda_i)]}{\sum_{i=1}^n \text{sen}^2[k\pi \cos(\lambda_i)]} \left(\frac{n-1}{n} \right). \quad (5.34)$$

Esta es una expresión que relaciona directamente los parámetros con los datos. Como se indicó anteriormente, el coeficiente a_1 (Ec. 5.35) da la aproximación de primer orden para la función residual (Ec. 5.29):

$$a_1 = \frac{\sum_{i=1}^n i \text{sen}[\pi \cos(\lambda_i)]}{\sum_{i=1}^n \text{sen}^2[\pi \cos(\lambda_i)]} \left(\frac{n-1}{n} \right). \quad (5.35)$$

Si los datos consisten en una población vectorial con una alineación neta perpendicular a la dirección preferida, Ec. 5.29 se parecerá a la curva azul del panel derecho de la Fig. 5.4 con el coeficiente a_1 tomando valores positivos. Esto se debe a que los datos presentan un exceso en los valores inferiores de los cosenos y, como consecuencia, la ECDF toma valores mayores que la función de control de manera que los residuos son positivos. Por otro lado, si los datos presentan un alineamiento neto en la dirección paralela, a_1 tomará valores negativos.

5.2. Aplicación a datos sintéticos, comparación con métodos tradicionales, y el parámetro ζ

Para comprobar la eficacia de los métodos presentados anteriormente con respecto a los métodos habituales, como el coseno medio, los aplicamos a 3 conjuntos de datos sintéticos. Estos datos se generan mediante el sorteo de puntos aleatorios en la superficie de un elipsoide tridimensional con ejes a , b y c , con varias excentricidades definidas de la manera habitual: $e^2 = 1 - c^2/a^2$, donde $c < a$ y $a = b$. Cabe destacar que al variar el eje c , estamos definiendo esta dirección vertical correspondiente al eje z en un sistema cartesiano como la dirección privilegiada de la simetría esférica. Elegimos establecer tres excentricidades diferentes para probar los métodos: $e^2 = 0.6$, 0.4 y 0 , pasando de alargada a isotrópica, respectivamente. De este modo, simulamos una población de vectores sin orientación preferente para el caso isotrópico, hasta una con una fuerte tendencia a la alineación para la mayor excentricidad.

En primer lugar, estudiamos la estabilidad de los estimadores con varios tamaños de muestra. La Fig. 5.5 muestra la media y la desviación estándar calculadas para 50 realizaciones aleatorias de muestras de tamaño N_{ran} . Para una muestra de tamaño $N \gtrsim 100$ la estimación de los parámetros a_1 y η es fiable, y para tamaños de muestra de más de $\sim 10^4$ el error relativo es lo suficientemente pequeño como para distinguir pequeñas variaciones en las excentricidades. Esto es prometedor en el sentido de que, para los estudios actuales y futuros en gran escala con grandes muestras, incluso un pequeño efecto de alineación sería detectable con estos métodos.

En la Fig. 5.6 probamos la significación estadística del parámetro η cuando se compara con la media de los cosenos, $\langle \cos(\lambda) \rangle$. Como se ha explicado anteriormente, generamos puntos aleatorios a lo largo de la superficie de tres elipsoides con valores de excentricidad de 0.6 , 0.4 y 0 . Estas poblaciones producen los tres histogramas de cosenos mostrados en el panel a), que es la forma habitual de estudiar y visualizar las alineaciones, donde incluimos los valores medios junto con la desviación estándar de la distribución. Notamos que la desviación estándar de la distribución del coseno es del orden de su media.

El panel c) muestra el logaritmo del parámetro β definido como el cociente de las componentes perpendicular y paralela del vector, siendo la dirección para-

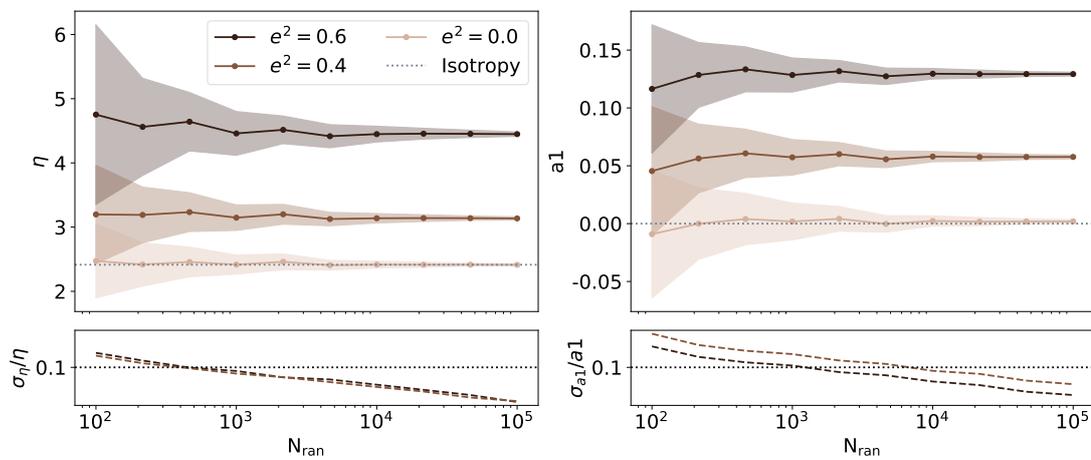


Figura 5.5: Estabilidad de los métodos con respecto al tamaño de la muestra. Aplicamos los métodos a 50 realizaciones aleatorias de datos sintéticos con alineaciones correspondientes a tres valores de excentricidades crecientes: 0, 0.4 y 0.6. La media y las desviaciones estándar, representadas por las líneas sólidas y las regiones sombreadas respectivamente, se calcularon con estos 50 resultados independientes. Encontramos que la media de los parámetros η y a_1 es estable incluso con un tamaño de muestra de unos pocos cientos. Se consigue un error relativo del 10% con un tamaño de muestra de $\sim 10^3$ para el parámetro η . El mismo tamaño de muestra para el mismo error relativo se consigue al aplicar el método de coeficiente OLS a los datos con la excentricidad más grande.

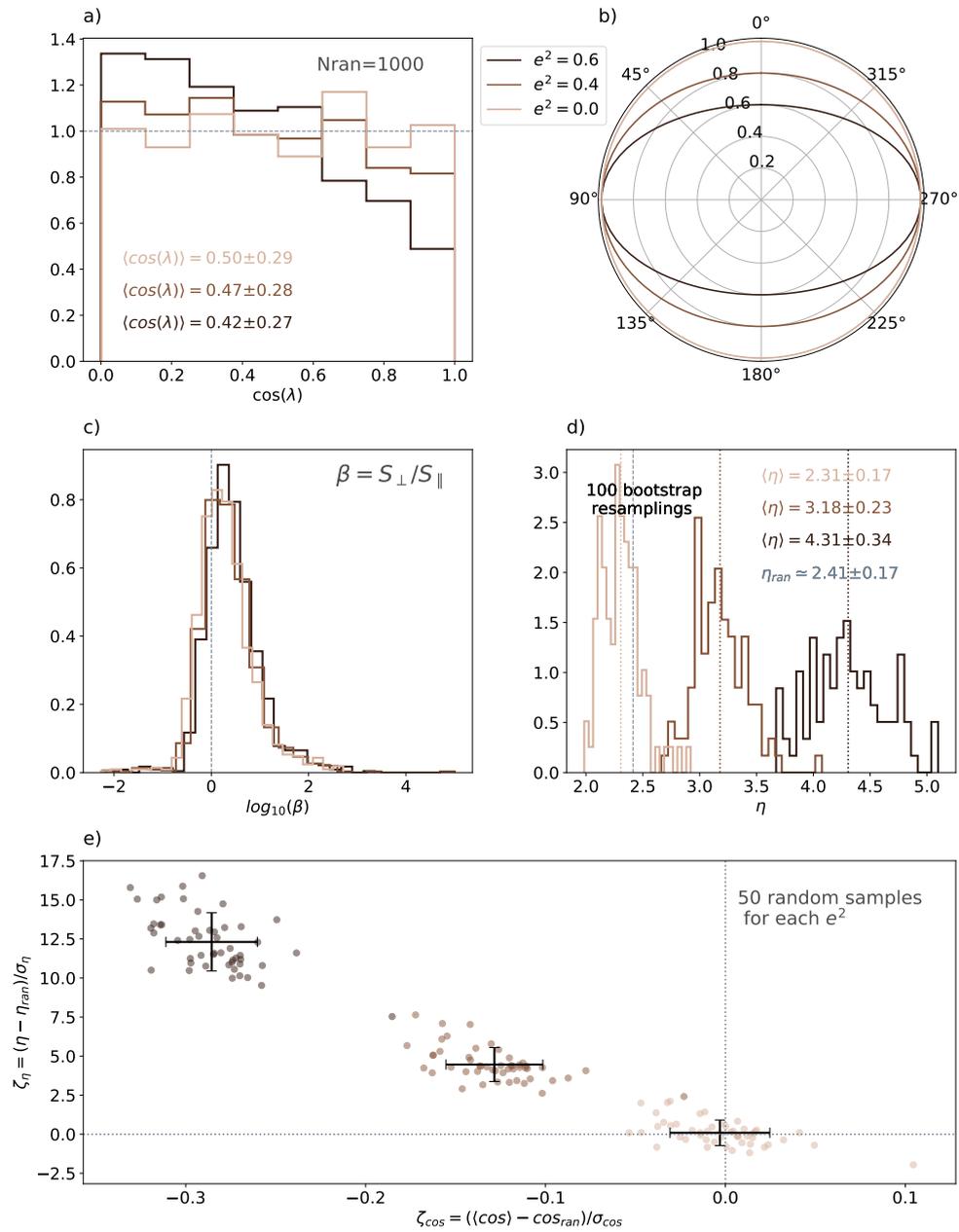


Figura 5.6: Representación del método a partir de una distribución de alineaciones de una población de vectores que llega al parámetro η de la “fracción de vectores”, junto con su significado comparado con la media de los cosenos.

lela la de la dirección privilegiada en la simetría esférica. El valor medio de β es similar para todas las poblaciones, por lo que no es un parámetro ideal para estudiar. Sin embargo, el número acumulado de vectores que tienen una componente perpendicular mayor, es decir, el parámetro η , es notablemente distinto para las diferentes muestras.

Para tener en cuenta la varianza de la muestra, realizamos un muestreo bootstrap de nuestro observable β con el fin de obtener una distribución de η a partir de la cual podemos definir una media y una desviación estándar. Estas distribuciones bootstrap para cada excentricidad se muestran en el panel d). Incluimos el valor medio y la desviación estándar de las distribuciones, así como la del caso isotrópico en color gris, que se ha calculado teóricamente con las expresiones detalladas en la Sec. 5.1.2.

Se observa que las distribuciones η tienen mayor desviación estándar para valores medios mayores. Esto es una consecuencia de la definición de β , donde los valores para el caso $S_{\perp} < S_{\parallel}$ están limitados entre 0 y 1, mientras que para el caso $S_{\perp} > S_{\parallel}$ no tienen un límite superior teórico. Si se definiera β a la inversa, se produciría el mismo comportamiento divergente para alineaciones paralelas. Esta es una característica del parámetro a tener en cuenta. Sin embargo, aunque el límite superior es infinito en teoría, no lo es en la práctica. Por un lado, habría que encontrar vectores de norma infinita para que esto suponga un inconveniente. Por otra parte, es poco probable encontrar la alineación correspondiente a una excentricidad de 0.6 en observables como las orientaciones galácticas, y mucho menos en valores de excentricidad más altos. En otras palabras, estamos probando estos parámetros en los límites extremos de situaciones prácticas.

Para evaluar finalmente la eficacia y significación del parámetro η , repetimos el procedimiento anterior generando puntos aleatorios a lo largo de la superficie de los diferentes elipsoides con 50 semillas aleatorias, obteniendo así 50 valores de η y cosenos medios. Además, para estudiar la significación estadística con respecto a los comportamientos aleatorios, definimos la variable ζ como

$$\zeta_X = \frac{X - \bar{X}_{ran}}{\sigma_{X,ran}}, \quad (5.36)$$

donde X es la variable aleatoria que queremos probar, en este caso: η y $\langle \cos(\lambda) \rangle$. Dada su definición, la variable ζ_X contiene información, no sólo sobre cuánto se desvía la variable X del comportamiento isotrópico, sino también sobre la importancia estadística de esta desviación. Los valores isotrópicos para la media y

la desviación estándar de η se han derivado teóricamente en la Sec. 5.1.2. La media y la desviación estándar de los cosenos en el caso isotrópico pueden calcularse como las de una distribución uniforme. La función de distribución de probabilidad de una distribución uniforme es:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{c.c.} \end{cases} \quad (5.37)$$

donde, en el caso de los cosenos, $a=0$ y $b=1$. Así que la media y la desviación estándar serían:

$$E(\text{COS}_{\text{ran}}) = 0.5, \quad (5.38)$$

y

$$\sigma_{\text{cos, ran}} = \sqrt{\frac{1}{12}} \simeq 0.289. \quad (5.39)$$

El panel e) muestra el cálculo de ζ_η y ζ_{cos} . Observamos que la significación estadística de las desviaciones del comportamiento isotrópico medidas con η es mucho mayor que con los cosenos. Para una excentricidad de $e^2 = 0.6$ tenemos una significación de alrededor de 12.5 para η y 0.3 para los cosenos.

En la Fig. 5.7 realizamos un análisis equivalente para el método de los coeficientes OLS utilizando los mismos datos sintéticos, como puede verse comparando las distribuciones del coseno en los paneles a) de ambas Figs. 5.7 y 5.6. Para este método calculamos primero la ECDF de los cosenos (panel c) de los datos correspondientes a los tres casos de anisotropía variable. Los residuos se calculan restando a la ECDF de los datos la correspondiente a una distribución isotrópica que es la recta de $ECDF(\cos(\lambda)) = \cos(\lambda)$. Generamos los datos y realizamos este cálculo 50 veces con diferentes semillas aleatorias, como se muestra en el panel d). Ajustamos cada una de estas curvas y representamos la media del ajuste con líneas sólidas y su correspondiente 3σ con las bandas sombreadas en el panel f).

La regresión lineal para cada curva produce un conjunto de coeficientes a_k , donde el que determina la forma básica del ajuste es a_1 (véase la sección 5.1.3). Realizamos un remuestreo bootstrap de los datos para estimar la media y la desviación estándar de este coeficiente. El panel e) muestra la distribución bootstrap del coeficiente a_1 correspondiente a las distribuciones del coseno mostradas en los mismos colores, donde las líneas verticales punteadas corresponden a la media. Además, en este panel indicamos en texto los valores medios junto con la des-

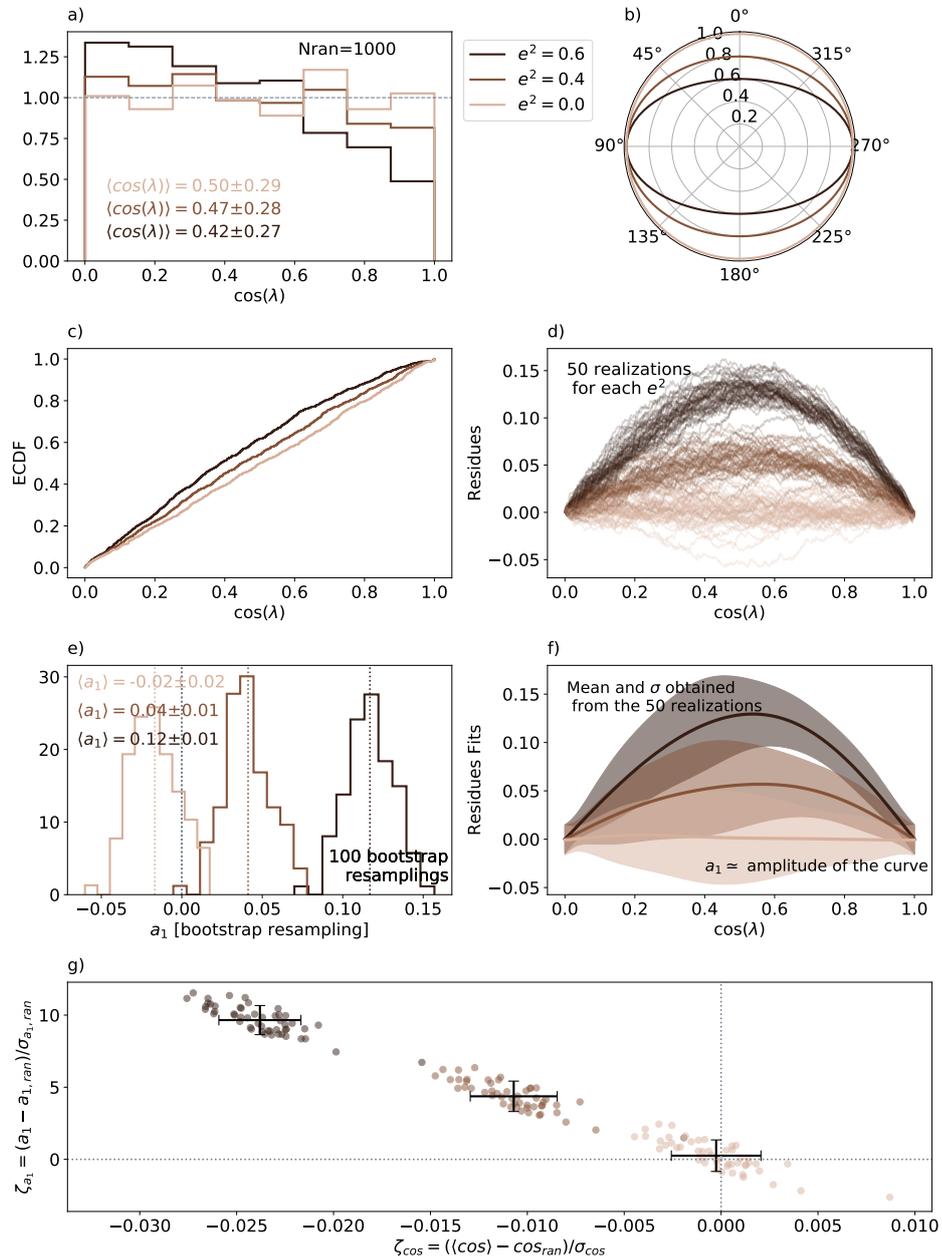


Figura 5.7: Representación del método partiendo de una distribución de alineamientos de una población de vectores y derivando el coeficiente OLS a_1 , junto con su significación en comparación con la media de los cosenos.

viación estándar de la distribución. Se observa fácilmente que la media de este coeficiente se determina de forma más robusta que la media de los cosenos.

Por último, para cada una de las realizaciones aleatorias trazamos los parámetros normalizados ζ_{a_1} y $\zeta_{(\cos)}$, donde encontramos que el coeficiente OLS detecta la alineación con mayor significación estadística.

5.3. Discusión y comentarios

En este trabajo presentamos dos métodos para detectar y cuantificar las señales de alineación y comprobar su robustez estadística. El primer método utiliza un parámetro denotado por la letra η , definido como la fracción relativa de componentes vectoriales en el plano perpendicular a una dirección de referencia. Hemos derivado el primer y segundo momento de la distribución de este estimador, y así podemos evaluar de forma fiable su significación estadística. El segundo método se basa en un ajuste sobre los residuos de la ECDF de los datos con respecto a la esperada para una distribución uniforme. El ajuste utiliza un pequeño conjunto de funciones armónicas ortogonales y no se basa en ningún esquema de discretización. La amplitud del ajuste, i.e., la amplitud de la señal de alineación, puede ser descrita por el primer parámetro OLS impar, a_1 .

Para el primer método, derivamos la distribución del estadístico de prueba β , definido como el cociente entre los componentes de los vectores perpendicular y paralelo (Ec. 5.4). Encontramos que la distribución de probabilidad $f_B(\beta)$ es una distribución patológica donde los momentos son indefinidos, y como tal, no es un estadístico robusto. Sin embargo, utilizando este estadístico consideramos la fracción de sus valores que es mayor que uno, dado que cuando las componentes perpendicular y paralela son iguales no hay preferencia en ninguna de las dos direcciones y entonces $\beta = 1$ puede tomarse como valor crítico. Por lo tanto, definimos este cociente como el parámetro η en la Ec. 5.15. Encontramos que, para β definida como en la Ec. 5.4, el parámetro η tiene un valor de expectación y una varianza dadas por $\eta_0 \simeq 2.4142$ y $\text{Var}(\eta) \simeq 28.1421/N$, respectivamente. El comportamiento gaussiano de este parámetro permite que los dos primeros momentos sean suficientes para describir su distribución. La ventaja de conocer la distribución teórica de los parámetros es la capacidad de determinar con precisión la importancia estadística de cualquier señal detectada sin invertir tiempo y recursos computacionales en simulaciones de Monte Carlo.

El segundo método de análisis de alineación que presentamos arroja los coeficientes OLS de un ajuste de los residuos de la ECDF de los cosenos de los datos con respecto a una muestra aleatoria. El primer coeficiente impar, a_1 , de la expansión armónica de la función residual es suficiente para caracterizar la amplitud de la señal de alineación, siendo cero cuando la muestra es consistente con orientaciones isotrópicas. Los valores positivos de a_1 indican una alineación perpendicular, mientras que los valores negativos indican una alineación paralela con respecto a la dirección privilegiada.

Hemos comparado estos métodos con la media de la distribución del coseno y utilizando simulaciones de Monte Carlo. Esta comparación se ha realizado generando puntos aleatorios en la superficie de elipsoides con distintas excentricidades, simulando así una población de vectores con tres niveles diferentes de alineación (desde la no alineación, a la intermedia, a la muy alineada), y probando qué tan fidedigna es la recuperación de la señal de alineación tanto con los parámetros nuevos como con los tradicionales.

Encontramos que los estadísticos propuestos permiten la detección de señales de alineación con una mayor significación. Para una desviación de aproximadamente $0,25\sigma$ de una distribución isotrópica de cosenos, obtenemos una significación de $10-$ y 12σ para el coeficiente OLS a_1 y el parámetro η respectivamente. En el capítulo siguiente comentamos los resultados de [Dávila-Kurbán et al. \(2022\)](#) donde aplicamos el primer método presentado en este capítulo, es decir, el parámetro de fracción de vectores alineados, η , a los datos de una simulación cosmológica.

Por último, hemos evaluado los efectos de las incertidumbres en la medición de las componentes vectoriales paralelas y perpendiculares. Para ello, modelamos los errores de observación introduciendo ruido gaussiano en las componentes. La desviación estándar se eligió como el 10% de la norma vectorial media. Existe una relación lineal entre los parámetros “reales”, η y a_1 , obtenidos con los datos sintéticos brutos, y los parámetros “observados” que tienen en cuenta el error de observación simulado. Realizamos 100 realizaciones de los cálculos con y sin errores simulados de observación. La ordenada al origen b correspondiente a la máxima divergencia de los valores reales puede expresarse, para un tamaño de muestra N de 1000 y 5000 respectivamente, como $\Delta b_\eta = 0.7$ y 0.3 para el primer método, y $\Delta b_{a_1} = 0.03$ y 0.01 para el segundo método. Como referencia, el rango de valores de los dos parámetros para las tres excentricidades probadas son:

$2.0 < \eta < 5.5$ y $-0,05 < a_1 < 0,16$ para $N=1000$; y $2,2 < \eta < 5,0$ y $-0,02 < a_1 < 0,15$ para $N=5000$. Esto puede utilizarse para estimar aproximadamente el tamaño de las muestras necesarias para lograr la detección de una pequeña señal en los datos. Con estas herramientas, los próximos relevamientos y simulaciones de gran tamaño pueden proporcionar nuevos conocimientos al detectar pequeñas señales de los alineamientos que no se detectan en los relevamientos actuales con un número menor de galaxias.

Capítulo 6

Alineaciones de galaxias en voids

En este capítulo, basado en el trabajo [Dávila-Kurbán et al. \(2022\)](#), estudiamos la alineación de los espines galácticos en las cáscaras de los vacíos con respecto a la dirección radial utilizando un identificador de vacíos bien establecido ([Ruiz et al. 2015](#)) y analizamos la dependencia de la amplitud de la señal con la masa de la galaxia, la norma de espín, la velocidad y la densidad local de las mismas. Al incluir las velocidades en nuestro análisis de alineación de espín, estudiamos un aspecto dinámico que no ha sido muy explorado antes. Por último, empleamos un método novedoso que utiliza parámetros estadísticos estables y robustos para rechazar o aceptar la hipótesis nula de no alineación ([Dávila-Kurbán et al., 2022](#)). Es de esperar que este nuevo enfoque proporcione una perspectiva útil sobre el estudio de las orientaciones de las galaxias.

La simetría esférica de los vacíos, tanto en su geometría como en su dinámica, permite definir una dirección específica para analizar las orientaciones galácticas: la dirección radial. Dado el problema de las orientaciones vectoriales en torno a un punto central definimos los parámetros β , η y ζ que nos permitirán estudiar la orientación de las galaxias y detectar posibles desvíos respecto a una distribución aleatoria. Estos parámetros se introdujeron y analizaron formalmente en el [Cap. 5](#). La [Fig. 6.1](#) muestra un resumen esquemático de las definiciones de los parámetros, cómo se relacionan entre sí y, en definitiva, cómo partimos de la medición de un ángulo hasta la representación visual de la señal de alineación que utilizamos para mostrar nuestros resultados.

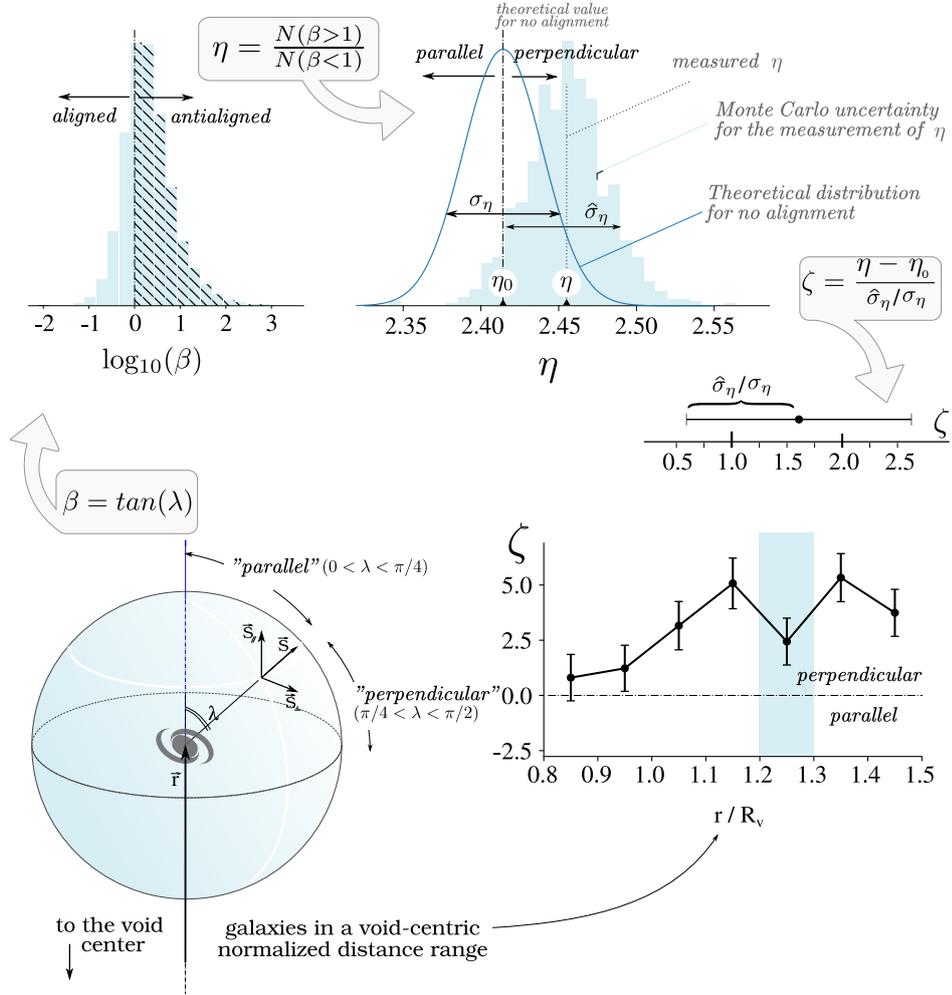


Figura 6.1: Esquema del análisis realizado en la búsqueda de señales de alineación de discos de galaxias alrededor de vacíos. Comenzamos con un parámetro λ , el ángulo agudo entre el vector de espín y la posición del vector de la galaxia con respecto al centro del vacío. A continuación, se compara la proporción de galaxias con $\tan(\lambda) > 1$ y con $\tan(\lambda) < 1$ con una distribución teórica para una muestra de galaxias orientadas al azar para diferentes intervalos de distancia radial al centro del vacío. Finalmente, se define un parámetro normalizado, ζ , para mostrar tanto la tendencia de alineación, si la hay, como su importancia estadística; por ejemplo, si $\zeta > 3$, entonces esta población de galaxias muestra una tendencia a estar alineada perpendicularmente con una confianza de más de 3σ .

6.1. Datos: TNG300-1

En este trabajo empleamos datos de galaxias del proyecto Illustris-TNG (TNG, Marinacci et al. 2018, Naiman et al. 2018, Nelson et al. 2018, 2019, Pillepich et al. 2019, 2018, Springel et al. 2018). Illustris-TNG es un conjunto de simulaciones magneto-hidrodinámicas cosmológicas obtenidas con el código de malla móvil AREPO (Springel 2010), y adoptando la cosmología de Planck (Planck Collaboration et al. 2016): $\Omega_m = 0.3089$, $\Omega_b = 0.0486$, $\Omega_\Lambda = 0.6911$, $\sigma_8 = 0.8159$, $n_s = 0.9667$, y $h = 0.6774$. Estas simulaciones presentan modelos exhaustivos para la física de la formación de galaxias, y mejoran a su predecesor, Illustris, al incluir campos magnéticos y mejorar los modelos del viento galáctico y la retroalimentación de AGN. El proyecto TNG abarca tres volúmenes diferentes con condiciones iniciales y modelos físicos idénticos: TNG50, TNG100 y TNG300. En particular, empleamos el TNG300-1, con una caja periódica de $205h^{-1}\text{Mpc}$, la caja más grande y de mayor resolución del conjunto. Los halos (grupos) y subhalos (galaxias) en TNG se encuentran con un algoritmo estándar de “amigos de los amigos” (FoF, por sus siglas en inglés *friends of friends*) con una longitud de enlace $b = 0.2$ (en unidades del espaciado medio entre partículas) ejecutado en las partículas de materia oscura, y con el algoritmo SUBFIND (Springel et al. 2001) respectivamente. Este último detecta la subestructura dentro de los grupos y define grupos de partículas localmente sobredensos y auto-ligados, donde el componente bariónico en la subestructura se define como una galaxia. Analizamos esta simulación a redshift $z = 0$ considerando galaxias en un el rango de masa $10^9 M_\odot \leq M_\star \leq 10^{13} M_\odot$. El espín de las galaxias en la suite TNG se define como el espín total por eje calculado como la suma ponderada por la masa de la coordenada relativa por la velocidad relativa de todas las partículas integrantes. El corte inferior en la masa mencionado anteriormente nos permite emplear sólo galaxias en las que el espín está bien definido.

6.2. Identificación de vacíos y su población de galaxias

La identificación de vacíos en la simulación sigue el algoritmo descrito en Ruiz et al. (2015), una versión modificada de algoritmos anteriores presentados en Padilla et al. (2005) y Ceccarelli et al. (2006). El algoritmo estima el perfil

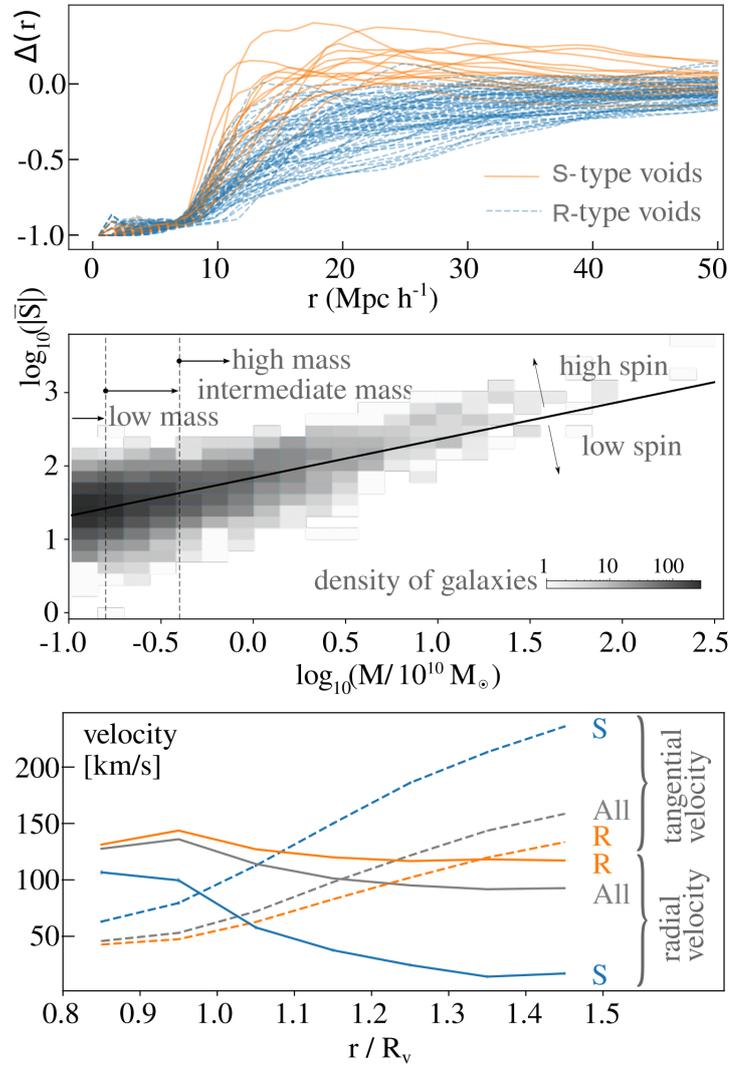


Figura 6.2: Esta figura resume varios aspectos de los datos con los que trabajamos. El panel superior muestra los perfiles de densidad de los vacíos que identificamos en la simulación, y su clasificación en vacíos tipo R y tipo S. El panel central representa el logaritmo de la norma del vector de espín \bar{S} de una galaxia en función del logaritmo de su masa en unidades de $10^{10} M_{\odot}$. Por último, el panel inferior muestra la velocidad radial y transversal en función de la distancia al centro del vacío en unidades de radio del vacío. Una velocidad radial positiva indica que los entornos que estudiamos están en expansión, especialmente los vacíos de tipo R, y por lo tanto deben ser entornos subdensos.

de densidad con una teselación de Voronoi sobre trazadores de densidad, que en este trabajo son galaxias de la simulación TNG300-1. Las regiones subdensas se obtienen seleccionando celdas de Voronoi por debajo de un umbral de densidad y se seleccionan como candidatas a vacío. Centrado en estas celdas, se calcula el contraste de densidad integrado $\Delta(r)$ en valores crecientes de r . Los candidatos a vacío se seleccionan como las esferas más grandes que satisfacen la condición $\Delta(R_v) < 0.9$ donde R_v es el radio del vacío. A continuación, los centros de los vacíos se desplazan aleatoriamente para que las esferas puedan crecer. Esto se hace porque es probable que el algoritmo produzca vacíos esféricos en los que sus cáscaras no se ajustan con precisión a las estructuras circundantes, y el procedimiento de recentrado proporciona estructuras con bordes que concuerdan mejor con el campo de densidad local circundante. Por último, el catálogo de vacíos comprende las mayores esferas subdensas y no superpuestas de radio R_v . En este trabajo, usamos como trazadores a las galaxias de TNG300-1 con un corte inferior en masa de $M_{\min} = 10^9 M_{\odot}$ para evitar posibles problemas de resolución en la definición de las mismas. Después de aplicar este algoritmo a TNG300-1 y de cortar los vacíos con *shot-noise*¹, nos queda una muestra de 82 vacíos con radios en el rango de 7-11 $h^{-1}\text{Mpc}$.

Los entornos de los vacíos pueden proporcionar una visión física sobre la naturaleza y la evolución de las propiedades de los vacíos, ya que su jerarquía se deriva de la agrupación de la masa en la estructura creciente cercana (Paranjape et al. 2012, Sheth et al. 2004). Algunos vacíos colapsan sobre sí mismos con su estructura circundante, mientras que otros vacíos permanecen como regiones subdensas. Estos dos tipos de evolución están determinados por la densidad circundante: los vacíos rodeados por un entorno parecido a la densidad media de fondo se expandirán y seguirán siendo una región subdensa, lo que se conoce como un tipo de vacío-en-vacío, mientras que si el vacío está rodeado por una cáscara sobredensa, un sistema de vacío-en-nube, probablemente se encogerá bajo el colapso de la cáscara. Estos dos comportamientos evolutivos pueden identificarse calculando los perfiles de densidad radial acumulada (Ceccarelli et al. 2013, Paz et al. 2013, Ruiz et al. 2015). Empleamos subhalos como trazadores de la densidad alrededor de los vacíos; los mismos trazadores usados para identificar los mismos. En el panel superior de la Fig. 6.2 se muestra, en líneas anaranjadas sólidas, el contraste de densidad de los vacíos *shell* (“cáscara” en inglés), o tipo S, mientras

¹Ruido en pequeñas escalas característico de muestras discretas.

que las líneas azules discontinuas representan el de los voids *rising* (“ascendente” en inglés), o tipo R. En la práctica, esta clasificación se hace evaluando $\Delta(3R_v)$, tal que cuando este es positivo se clasifica al vacío como tipo R y como tipo S en el caso contrario. Estudiamos la señal de alineación en estos dos tipos de vacíos, así como en la muestra completa de vacíos.

6.2.1. Velocidades y entorno de las galaxias

En esta sección estudiamos la dinámica de las galaxias caracterizadas en la subsección anterior y su entorno local para interpretar mejor los resultados. Para ello, exploramos la velocidad radial y transversal media de las galaxias en función de su distancia al centro del vacío (tanto de tipo R como S), como se muestra en el panel inferior de la Fig. 6.2. Como es de esperar, a partir de aproximadamente $1R_v$, las velocidades radiales centradas en el vacío comienzan a disminuir, mientras que las velocidades transversales siguen aumentando impulsadas por las estructuras sobredensas, como los filamentos y los cúmulos masivos.

Destacamos el hecho de que en el rango analizado de distancias al centro de los vacíos, las velocidades radiales de las galaxias siguen siendo considerablemente grandes. Las medias de la velocidades radiales son de aproximadamente 25 y 120km s^{-1} , para galaxias en vacíos tipo S y R respectivamente, aún cuando se encuentran a $1.5R_v$. Esto confirma que estas regiones aún se encuentran en expansión global, asociada a un entorno subdenso en gran escala.

Para confirmarlo, calculamos el contraste de densidad acumulado, igual que en el panel superior de la Fig. 6.2, pero en función de los radios de los vacíos en el rango de distancias de $0.8-1.5R_v$. El contraste de densidad de los vacíos de tipo R está por debajo de -0.25 para todo este rango de distancias, y sólo para unos pocos vacíos supera el valor de -0.50 en las distancias más lejanas de $\sim 1.5R_v$.

6.2.2. Propiedades de las galaxias en los vacíos y su clasificación

En esta subsección nos centramos en explorar la dependencia con la alineación de las diferentes características de las galaxias. Consideramos las propiedades intrínsecas, como la masa y el espín total, y también su entorno de densidad local y la velocidad de expansión con respecto al centro del vacío. Para estudiar esto último, consideramos cáscaras cada vez más grandes y no superpuestas de un an-

cho de $0.1R_v$. Además, dividimos la población dentro de las cáscaras en muestras de “alta” y “baja” velocidad con respecto a los valores medios. Tenemos en cuenta la densidad local utilizando el parámetro estadístico Σ_5 definido en la Sec. 6.3.3. Encontramos que la velocidad de expansión y la densidad local son variables independientes, por lo que es factible estudiarlas por separado. Sin embargo, para distinguir las poblaciones de galaxias en alto y bajo espín es necesario analizar la correlación masa–espín.

El panel central de la Fig. 6.2 muestra el espín y la masa de una población de galaxias correspondiente a una cáscara de radios interiores y exteriores de 1.0 – $1.1R_v$ de un vacío con radio $R_v \simeq 8h^{-1}\text{Mpc}$. Una simple diferenciación en galaxias de alto/bajo espín se realiza mediante una regresión lineal en la relación espín–masa como se muestra con la línea sólida en el panel central de la Fig. 6.2 que divide la muestra en un número similar de objetos. Hemos considerado tres rangos de masa que corresponden aproximadamente a los terciles de la muestra: $M_1 = 10^{9.2}M_\odot$ y $M_2 = 10^{9.6}M_\odot$; estos dos límites se muestran en líneas discontinuas verticales.

El número total de galaxias en cáscaras de 0.8 – $1.5R_v$ alrededor de los 82 vacíos identificados es $N_{\text{Total}} = 413864$. Los números de cada uno de los siete intervalos de radios, o, de forma equivalente, siete cáscaras con una profundidad de $0.1R_v$ alrededor de los vacíos tipo R y S son, respectivamente: $N_R = [5666, 8855, 29273, 44152, 60364, 75188, 93348]$ y $N_S = [1015, 1872, 6770, 12946, 18412, 25741, 30262]$.

6.3. Efectos de alineaciones en las galaxias

En esta sección, exploramos la señal de alineaciones de los discos galácticos y su relación con las propiedades galácticas y características ambientales, como la norma de espín, la masa, la velocidad y la densidad local de las galaxias. Para ello, hacemos uso del parámetro ζ (Ec. 5.36), definido como normalización para el parámetro η (Ec. 5.15).

Nuestro análisis pretende determinar qué propiedades de la muestra de galaxias producen un cambio significativo en la señal de alineación. Para estudiar la dependencia de la señal de alineación ζ con respecto a la posición radial r , primero consideramos las galaxias pertenecientes a la muestra completa de vacíos, y luego consideramos las muestras de vacíos tipo R y tipo S por separado. Seleccionamos galaxias dentro de cáscaras esféricas de $0.1R_v$ de profundidad, de $0.8R_v$ a $1.5R_v$,

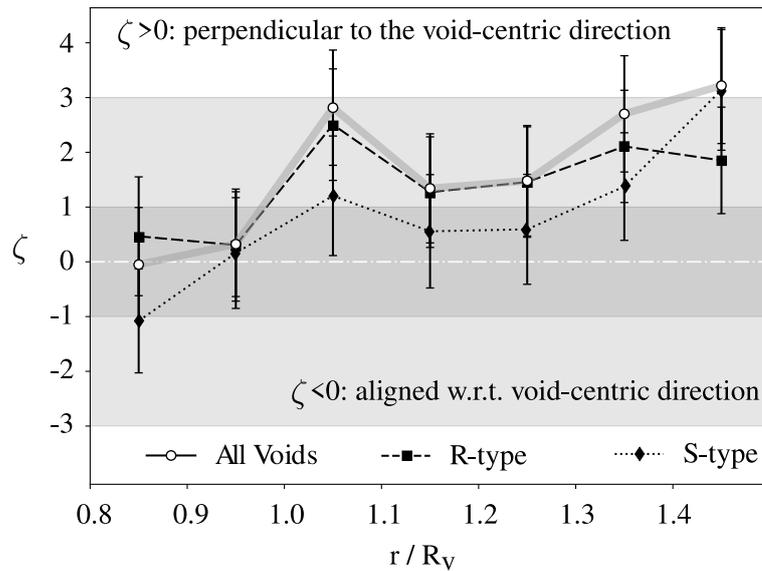


Figura 6.3: Señal de alineación en función de la distancia al centro del vacío en unidades de radio del vacío, $\zeta(r R_v^{-1})$. Las líneas discontinuas largas y cortas representan la señal alrededor de los vacíos de tipo R y S, respectivamente, mientras que la línea sólida representa los resultados de todos los tipos de vacíos. Hay un pico notable de señal perpendicular, o $\zeta > 0$, en la cáscara centrada en $1.05R_v$. Más lejos de este pico, la señal de los vacíos de tipo S, aunque en general es menor que la de los vacíos de tipo R, aumenta, superando a estos últimos en el intervalo de distancia más lejano. Aunque nada demasiado significativo, es decir, ninguna señal por encima de 3σ de confianza, existe una tendencia general de alineación perpendicular para toda la muestra de galaxias.

comprendiendo un total de 7 intervalos.

Una vez que hemos determinado la población de galaxias a estudiar, la dividimos en valores “bajos” y “altos” de la propiedad galáctica que estamos estudiando, por ejemplo bajo y alto espín, con la excepción de la masa que también tiene una clasificación “intermedia”, como se discute en la Sec. 6.2.2. A continuación, para cada una de las submuestras calculamos $\zeta(r R_v^{-1})$, es decir, la alineación en función de la posición radial con respecto al centro del vacío, y analizamos si cambia significativamente para las diferentes submuestras. Con este enfoque, buscamos tendencias de alineación paralelas ($\zeta < 0$) o perpendiculares ($\zeta > 0$). Primero mostramos la señal de alineamiento sin realizar ninguna clasificación en las galaxias, i.e. la señal de toda la población de galaxias en cada intervalo de distancia al centro del vacío (Fig. 6.3). Luego presentamos los resultados de las submuestras determinadas por las clasificaciones generales mencionadas anteriormente (Fig. 6.4 y Secs. 6.3.1 a 6.3.4). Finalmente, en la Sec. 6.3.5 estudiamos el alineamiento en todas las posibles combinaciones de los mencionados criterios de clasificación.

En la Fig. 6.3 mostramos la señal de alineación, cuantificada por los valores promediados del parámetro ζ en función de la distancia al centro del vacío, considerando la muestra completa de vacíos, junto con las submuestras de tipo R y S. Las regiones de color gris claro y oscuro representan la significación de 3σ y 1σ , respectivamente, calculada con la Ec. 5.23, mientras que las barras de error se calcularon con 1000 remuestreos bootstrap de la población de galaxias en cada intervalo de distancia a lo largo de todos los vacíos en consideración. Como se puede observar, existe una tendencia general que favorece una señal de alineación perpendicular de los vectores de espín con respecto a la dirección radial. Este resultado puede interpretarse como indicativo de una preferencia de los discos a encontrarse perpendicularmente al centro del vacío (es decir, $\zeta > 0$). La confianza estadística de este exceso de probabilidad está entre 1 y 3σ para toda la muestra en distancias entre 1 y 1.5 radios de vacío. A pesar de ser una señal marginal, es consistentemente positiva en un amplio rango, cubriendo al menos cinco intervalos en la distancia normalizada. En general, la señal de los vacíos de tipo R parece ser más fuerte que la de los vacíos de tipo S, y hay un pico de alineación notable para ambos tipos en la cáscara centrada en $1.05R_v$. En los intervalos más alejados, la señal de los vacíos tipo S aumenta más rápidamente que la de los vacíos tipo R.

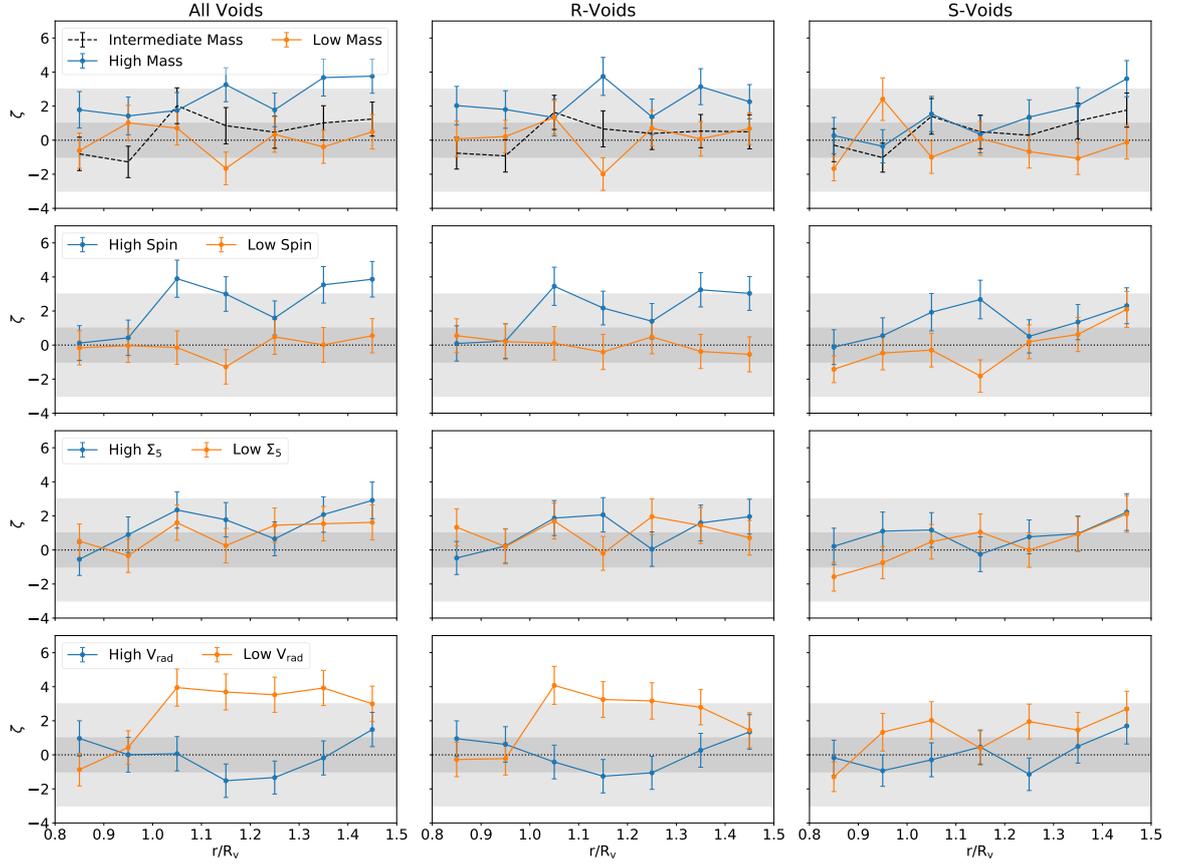


Figura 6.4: Resultados de la señal de alineación en función de la distancia radial al centro del vacío, $\zeta(r R_v^{-1})$, para diferentes muestras de galaxias. Las líneas azules y naranjas representan las muestras con valores altos y bajos de la propiedad galáctica filtrante, respectivamente. Hemos incluido el rango de masa intermedio en la primera fila, indicado con una línea negra punteada. Las regiones fuertemente sombreadas abarcan un nivel de confianza de 1σ mientras que el sombreado claro representa regiones de confianza de 3σ . Estas regiones corresponden a las incertidumbres de las muestras de referencia y se calculan con la expresión teórica de la derivación de ζ , mientras que las barras de error de la señal se obtienen del remuestreo bootstrap. Las señales de más de 3σ se encuentran al filtrar para masa alta, espín alto y velocidad radial baja. El filtrado por alto y bajo Σ_5 no parece tener un efecto significativo. La propiedad galáctica que produce la señal más fuerte en este rango de distancias radiales es la baja velocidad radial.

En la Fig. 6.4 mostramos los resultados de dividir la muestra de galaxias según las clasificaciones alta y baja de las propiedades que analizamos: masa, espín, densidad local y velocidad. Las líneas azul y naranja representan poblaciones con valores altos y bajos del parámetro de interés, respectivamente. Exploraremos los resultados de la división de la muestra con respecto a los parámetros mencionados en las siguientes subsecciones.

6.3.1. Alineación vs. masa

Seguimos la clasificación de masas discutida en la Sec. 6.2.2, y mostramos en la primera fila de la Fig. 6.4 los resultados de la señal de alineación para muestras con masa baja, intermedia y alta. Las muestras de masa baja e intermedia son en su mayoría consistentes con alineaciones isotrópicas. Sin embargo, se encuentra una señal consistentemente por encima de 1σ cuando se analizan muestras de alta masa. En particular, las cáscaras con una señal perpendicular significativa, es decir, por encima de la región sombreada de 3σ , se encuentran centradas en $1.15R_v$ alrededor de los vacíos de tipo R, y en $1.45R_v$ alrededor de los vacíos de tipo S. Teniendo en cuenta todos los tipos de vacíos, también se encuentra una señal más fuerte en las cáscaras centradas en $1.35R_v$.

Estos resultados no indican un claro “spin-flip” (ver Sec. 2.3) dependiente de la masa en las cáscaras de los vacíos con esta clasificación en masas. Sin embargo, los resultados sugieren una tendencia a pasar de la no alineación a la alineación perpendicular con el aumento de la masa. Hay un fenómeno que se asemeja a un *spin-flip* para las cáscaras centradas en $1.15R_v$ alrededor de los vacíos de tipo R, en los que las galaxias de baja masa parecen mostrar una señal de alineación de $\zeta \simeq -2 \pm 1$. Sin embargo no es una señal lo suficientemente fuerte como para que concluyamos a favor de la existencia de un *spin-flip* en este caso.

6.3.2. Alineación vs. espín

Seguimos la clasificación de espín discutida en la Sec. 6.2.2, y mostramos en la segunda fila de la Fig. 6.4 los resultados de la señal de alineación para las muestras de bajo y alto espín.

Los resultados de las galaxias de bajo espín son en su mayoría consistentes con la ausencia de alineación en todas las clasificaciones de los vacíos, es decir, $\zeta(r R_v^{-1}) \simeq 0$. Por otro lado, las galaxias de alto espín, es decir, las galaxias que

han adquirido una rotación relevante, muestran una fuerte y significativa ($\zeta \geq 3$) tendencia a ser perpendiculares para $r > 1R_v$, especialmente en los vacíos de tipo R. Las cáscaras que muestran una señal superior a 3σ se centran en torno a 1.05, 1.15, 1.35 y $1.45R_v$ en la muestra de todos los tipos de vacíos, y $1.05R_v$ en los vacíos de tipo R. En los vacíos de tipo S no se encuentra esta señal. Además, para los vacíos de tipo S no parece haber una diferencia estadísticamente significativa en la alineación entre las galaxias de bajo y alto espín para $r > 1.2R_v$.

6.3.3. Alineación vs. densidad del entorno local

El enfoque de “vecinos cercanos” estudia la densidad del entorno considerando un estimador de escala variable. Normalmente el parámetro de densidad superficial se calcula como $\Sigma_n = n/\pi r_n^2$, donde n es el número de vecinos dentro de una circunferencia con radio igual a r_n , la distancia al n -ésimo vecino más cercano. Definido de este modo, las galaxias con vecinos más cercanos, es decir, con un Σ_n mayor, se encuentran en entornos más densos. En este trabajo elegimos utilizar Σ_5 , definido como

$$\Sigma_5 = \frac{5}{\pi r_5^2}. \quad (6.1)$$

Las medianas promedio de Σ_5 , es decir, los valores críticos por los que dividimos la muestra en “alta” y “baja”, a través de todos los intervalos de distancia son $\langle M(\Sigma_{5, total}) \rangle = 4.01e-6 \text{ Mpc}^{-2} h$, $\langle M(\Sigma_{5, R}) \rangle = 3.66e-6 \text{ Mpc}^{-2} h$, y $\langle M(\Sigma_{5, S}) \rangle = 5.58e-6 \text{ Mpc}^{-2} h$.

Se puede observar en la tercera fila de la Fig. 6.4 que no encontramos diferencias estadísticamente significativas al filtrar la muestra en valores altos y bajos de Σ_5 . La curva de alto Σ_5 para la muestra completa de vacíos parece ser cualitativamente similar a la curva análoga de la Fig. 6.3, mientras que la curva de bajo Σ_5 exhibe una amortiguación aún mayor de la señal. Esto significa que al seleccionar una densidad local alta o baja no estamos afectando a la detección de la señal de alineación más allá de diluirla debido a una reducción en el tamaño de la muestra. En otras palabras, la alineación parece ser independiente de la densidad local de galaxias.

6.3.4. Alineación vs. velocidad respecto al centro del vacío

En esta subsección exploramos la relación entre la orientación del espín y las velocidades vacío-céntricas de las galaxias, que podrían mantener reliquias de

la dirección de encuentro preferida y de la adquisición del espín. Definimos las velocidades radiales y transversales como

$$v_{\text{rad}} = \mathbf{v} \cdot \hat{\mathbf{r}} \quad (6.2)$$

$$v_{\text{tra}} = \sqrt{\mathbf{v}^2 - v_{\text{rad}}^2} \quad (6.3)$$

respectivamente, donde v es la velocidad total de la galaxia y $\hat{\mathbf{r}}$ su dirección radial.

Encontramos una diferencia significativa entre las muestras de velocidad radial baja y alta. La última fila de la Fig. 6.4 muestra que se encuentra una señal perpendicular mucho mayor para las galaxias con *baja* velocidad radial. La diferencia en la señal de alineación entre las muestras con baja y alta velocidad radial es particularmente fuerte en los vacíos tipo R. También cabe destacar que todos los tipos de vacíos la señal persiste por encima de la región de confianza de 3σ para cada cáscara de $r > 1R_v$. Por otra parte, también hemos explorado submuestras con velocidad transversal baja y alta encontrando señales de alineación similares en ambos casos (no presentadas en la figura), i.e., filtrar por velocidad transversal no parece afectar la señal de alineación. Por comparación de la última fila de la Fig. 6.4 con las anteriores se puede observar que la velocidad radial aparece como la propiedad galáctica que más fuertemente correlaciona con la señal de alineación de espín perpendicular a la dirección radial del vacío.

6.3.5. Alineación en combinaciones de muestras

Hemos explorado previamente la dependencia de la alineación de 5 propiedades galácticas por separado: norma de espín, masa, Σ_5 , velocidad transversal y radial, encontrando la mayor señal de alineación de espín para galaxias de baja velocidad radial de expansión respecto del centro del vacío. En esta subsección estudiamos las alineaciones de espín en combinaciones de submuestras considerando valores altos y bajos de los parámetros galácticos mencionados. Para simplificar las referencias cruzadas, denominamos a estas submuestras del uno al 80 como “Ss1” (submuestra uno) a “Ss80” (submuestra 80), y a toda la muestra de galaxias sin discriminar según ninguna propiedad galáctica la denominamos “S0” (muestra cero).

Para simplificar, también consideramos las galaxias dentro de un único cascarón con una profundidad de $0.5R_v$, con radios interiores y exteriores de 0.9 y

$1.4R_v$ respectivamente, dando un único valor del parámetro ζ para cada tipo de vacío.

Dado que en las secciones anteriores encontramos que la velocidad radial baja es la propiedad galáctica que más se correlaciona con la señal perpendicular, hemos dividido el conjunto total de resultados en dos tablas. La tabla 6.1 muestra todas las combinaciones posibles de propiedades galácticas altas y bajas restringidas fijando que tengan alta velocidad radial centrada en el vacío, mientras que el conjunto restante de submuestras con baja velocidad radial se muestra en la Tabla 6.2.

Las submuestras con señales de alineación por encima del nivel de confianza de 3σ , i.e. $|\zeta| > 3$, están resaltadas en negrita en las tablas 6.1 y 6.2. Se trata de S0 y submuestras con alta masa, alto espín o bajas velocidades radiales. Este resultado es coherente con los presentados en las subsecciones anteriores. Las señales de alineación de espín más fuertes se encuentran aproximadamente en el nivel de 9σ para las submuestras Ss55 y Ss57. Estas dos submuestras tienen en común altos valores de espín y una baja selección de velocidad radial centrada en el vacío (ver Tabla 6.2).

La restricción de galaxias de alto o bajo Σ_5 amortigua la señal-ruido de la muestra progenitora; por ejemplo, Ss9 y Ss18 con respecto a S0, al igual que Ss17 y Ss26 con respecto a Ss8. Esto confirma aún más nuestro hallazgo de que la densidad local, medida con el parámetro Σ_5 , no está directamente correlacionada con la alineación del espín alrededor de los vacíos.

Además, para estudiar la dependencia de la señal de alineamiento con las otras propiedades galácticas, podemos mirar las submuestras Ss1 (alto espín), Ss8 (alta masa) y Ss54 (baja velocidad), tenemos $\zeta = 5.8 \pm 1.0$ para alto espín y $\zeta = 5.6 \pm 1.1$ para alta masa, para todos los tipos de vacío, por lo que ninguno de estos parámetros se correlaciona más fuertemente con la alineación que el otro. Concluimos que la selección de galaxias de alta masa y alto espín tiene una fuerte incidencia en alineaciones sistemáticas de espín. Además, la señal más fuerte obtenida para estas tres submuestras y para todos los tipos de vacíos, es Ss54 a $\zeta = 6.8 \pm 1.0$. Lo que confirma nuestro hallazgo de que, entre los parámetros analizados, una baja velocidad radial respecto al centro del vacío es el mayor predictor de la alineación.

En cuanto a la clasificación de los vacíos, encontramos valores de señal más significativos desde el punto de vista estadístico alrededor de los vacíos de tipo

R que alrededor de los vacíos de tipo S. El valor más alto para los vacíos de tipo R se encuentra en Ss55 (alto espín y baja velocidad radial) con $\zeta_R = 8.0 \pm 1.1$ mientras que el valor más alto para los vacíos de tipo S pertenece a Ss57 (alta masa y espín y baja velocidad radial) con $\zeta_S = 5.7 \pm 1.2$. Ambas submuestras son bajas en velocidad radial y pueden verse en la Tabla 6.2. Como se ha visto en la Sec. 6.3.4, para una determinada distancia al centro del vacío, los vacíos de tipo R habitan un entorno menos denso que los vacíos de tipo S. Esto significa que detectamos una mayor señal de alineación en entornos globalmente menos densos.

Elegimos la submuestra con la mayor señal, Ss57 (alto espín y masa, y baja velocidad radial), para trazar su señal de alineación normalizada ζ en función de la distancia al centro del vacío en el panel superior de la Fig. 6.5. Encontramos una gran señal de más de 5σ para la alineación perpendicular en torno a $1.15R_v$ en los vacíos de tipo R. El panel inferior muestra el contraste de densidad acumulada en las distancias consideradas, y muestra que las grandes señales se encuentran efectivamente en entornos subdensos.

La relación entre los parámetros de alineación analizados aquí y las mediciones más clásicas, como el coseno medio de la muestra, se describe en la Sec. 5.2. Para Ss57, por ejemplo, tenemos $\zeta_{\text{Total}} \simeq 9.0$ con $N_{\text{Total}} = 23853$ (ver Tabla 6.2); siguiendo la relación expuesta en la sección mencionada podemos estimar que este valor de ζ corresponde a un valor de coseno medio de $\langle \cos(\theta) \rangle \simeq 0.43$.

6.4. Discusión y comentarios

En este capítulo hemos analizado las orientaciones de los espines galácticos en entornos subdensos en la simulación Illustris TNG300-1. Nuestro estudio muestra una fuerte evidencia de que las grandes galaxias en vacíos cósmicos exhiben un exceso de espines perpendiculares a la dirección radial céntrica del vacío. La estadística utilizada para detectar la señal de alineación es robusta y permite explorar su dependencia con diferentes regiones en el espacio de parámetros, incluyendo la distancia radial al centro del vacío, la magnitud del espín, la masa de la galaxia, la densidad local de la galaxia y la componente radial de la velocidad de las galaxias respecto al centro del vacío cósmico.

Encontramos la mayor señal de alineación (a un nivel superior a 9σ) para galaxias masivas con rotación relevante que residen en entornos de vacío y con

Submuestra	Spin	Masa	Σ_5	V_{rad}	ζ_{Total}	ζ_R	ζ_S	N_{Total}	N_R	N_S
S0	-	-	-	-	4.0 ±1.0	3.6 ±1.0	1.8 ±1.0	283573	217832	65741
Ss1	H	-	-	-	5.8 ±1.0	5.0 ±1.0	3.0 ±1.1	156192	120793	35399
Ss2	H	L	-	-	-0.1 ±1.0	0.5 ±1.0	-1.1 ±1.0	50217	39109	11108
Ss3	H	H	-	-	8.0 ±1.1	6.6 ±1.1	4.7 ±1.1	47707	36607	11100
Ss4	L	-	-	-	-0.4 ±1.0	-0.2 ±1.0	-0.5 ±1.0	127381	97039	30342
Ss5	L	L	-	-	-0.4 ±1.0	-0.3 ±1.0	-0.1 ±1.0	30580	23507	7073
Ss6	L	H	-	-	0.0 ±1.0	0.5 ±1.0	-0.8 ±1.0	43598	32963	10635
Ss7	-	L	-	-	-0.4 ±1.0	0.2 ±1.0	-1.0 ±0.9	80797	62616	18181
Ss8	-	H	-	-	5.6 ±1.1	5.0 ±1.0	2.6 ±1.0	91305	69570	21735
Ss9	-	-	H	-	3.3 ±1.1	2.5 ±1.0	2.0 ±1.1	141785	108916	32870
Ss10	H	-	H	-	4.4 ±1.0	3.5 ±1.0	3.3 ±1.1	78096	60396	17699
Ss11	H	L	H	-	-0.3 ±1.0	0.6 ±1.0	-1.9 ±0.9	25108	19554	5554
Ss12	H	H	H	-	4.8 ±1.1	3.1 ±1.1	4.1 ±1.1	23852	18303	5550
Ss13	L	-	H	-	0.0 ±1.0	0.0 ±1.0	-0.1 ±1.0	63690	48519	15171
Ss14	L	L	H	-	-0.4 ±1.0	-0.2 ±1.0	0.1 ±1.0	15290	11753	3536
Ss15	L	H	H	-	0.2 ±1.0	0.8 ±1.0	-0.1 ±1.0	21799	16481	5317
Ss16	-	L	H	-	-0.4 ±1.0	0.3 ±1.1	-1.4 ±0.9	40398	31308	9090
Ss17	-	H	H	-	3.5 ±1.1	2.9 ±1.0	2.8 ±1.1	45652	34785	10867
Ss18	-	-	L	-	2.3 ±1.0	2.5 ±1.1	0.6 ±1.0	141786	108916	32870
Ss19	H	-	L	-	3.8 ±1.0	3.6 ±1.1	0.9 ±1.0	78096	60396	17699
Ss20	H	L	L	-	0.1 ±1.0	0.0 ±1.0	0.4 ±1.0	25108	19554	5554
Ss21	H	H	L	-	6.5 ±1.1	6.2 ±1.2	2.6 ±1.1	23853	18303	5550
Ss22	L	-	L	-	-0.4 ±1.0	-0.1 ±1.0	-0.5 ±1.0	63690	48519	15171
Ss23	L	L	L	-	-0.2 ±1.0	-0.4 ±1.0	-0.2 ±1.0	15290	11753	3536
Ss24	L	H	L	-	-0.2 ±1.0	-0.1 ±1.0	-0.9 ±1.0	21799	16481	5317
Ss25	-	L	L	-	-0.1 ±1.0	-0.2 ±1.0	0.1 ±1.0	40398	31308	9090
Ss26	-	H	L	-	4.3 ±1.0	4.2 ±1.1	1.0 ±1.0	45652	34785	10867
Ss27	-	-	-	H	-1.1 ±1.0	-1.0 ±1.0	-0.3 ±1.0	141786	108916	32870
Ss28	H	-	-	H	-0.7 ±1.0	-0.8 ±1.0	-0.2 ±1.0	78096	60396	17699
Ss29	H	L	-	H	-1.9 ±1.0	-1.4 ±1.0	-1.5 ±0.9	25108	19554	5554
Ss30	H	H	-	H	2.5 ±1.0	2.5 ±1.1	1.1 ±1.0	23853	18303	5550
Ss31	L	-	-	H	-0.8 ±1.0	-0.8 ±1.0	-0.5 ±1.0	63690	48519	15171
Ss32	L	L	-	H	0.4 ±1.0	0.0 ±1.0	1.2 ±1.0	15290	11753	3536
Ss33	L	H	-	H	-0.8 ±1.0	-0.8 ±1.0	-0.4 ±1.0	21799	16481	5317
Ss34	-	L	-	H	-1.3 ±1.0	-1.2 ±1.0	-0.6 ±1.0	40398	31308	9090
Ss35	-	H	-	H	1.2 ±1.0	1.2 ±1.0	0.3 ±1.0	45652	34785	10867
Ss36	-	-	H	H	-0.4 ±1.0	-0.7 ±1.0	0.2 ±1.0	57201	45592	13017
Ss37	H	-	H	H	-0.7 ±1.0	-0.6 ±1.0	0.5 ±1.0	32361	25785	7196
Ss38	H	L	H	H	-1.2 ±1.0	-1.0 ±1.0	-2.0 ±0.9	10359	8325	2248
Ss39	H	H	H	H	1.0 ±1.0	0.8 ±1.0	1.9 ±1.1	9972	7905	2308
Ss40	L	-	H	H	-0.5 ±1.0	-0.5 ±1.0	-0.2 ±1.0	24819	19780	5822
Ss41	L	L	H	H	-0.6 ±1.0	-0.9 ±1.0	0.0 ±1.0	6056	4871	1357
Ss42	L	H	H	H	0.1 ±1.0	0.0 ±1.0	-0.6 ±0.9	8521	6713	2059
Ss43	-	L	H	H	-1.4 ±0.9	-1.3 ±1.0	-1.4 ±0.9	16425	13193	3599
Ss44	-	H	H	H	1.0 ±1.0	0.6 ±1.0	1.0 ±1.0	18507	14625	4343
Ss45	-	-	L	H	-1.0 ±1.0	-0.6 ±1.0	-0.5 ±1.0	84583	63324	19852
Ss46	H	-	L	H	-0.3 ±1.0	-0.5 ±1.0	-0.7 ±0.9	45735	34611	10502
Ss47	H	L	L	H	-1.4 ±0.9	-0.9 ±1.0	-0.2 ±1.0	14749	11228	3306
Ss48	H	H	L	H	2.5 ±1.1	2.6 ±1.1	0.0 ±1.0	13879	10398	3242
Ss49	L	-	L	H	-0.6 ±1.0	-0.6 ±1.0	-0.4 ±1.0	38870	28739	9349
Ss50	L	L	L	H	1.0 ±1.0	0.7 ±1.0	1.6 ±1.1	9234	6881	2179
Ss51	L	H	L	H	-1.1 ±1.0	-1.0 ±1.0	0.0 ±1.0	13278	9768	3258
Ss52	-	L	L	H	-0.5 ±1.0	-0.3 ±1.0	0.3 ±1.0	23973	18115	5491
Ss53	-	H	L	H	0.9 ±1.0	1.1 ±1.0	-0.3 ±1.0	27144	20160	6523

Cuadro 6.1: Esta tabla muestra los resultados de la medición de ζ para todos los tipos de vacíos, así como los vacíos de tipo R y S por separado para submuestras definidas como todas las combinaciones posibles de Spin, Masa y Σ_5 altos y bajos, y velocidad radial V_{rad} , excluyendo las muestras con V_{rad} bajos que se muestran en la Tabla 6.2. Los radios interior y exterior de la cáscara considerada son 0.9 y $1.4R_v$ respectivamente.

Submuestra	Spin	Masa	Σ_5	V_{rad}	ζ_{Total}	ζ_R	ζ_S	N_{Total}	N_R	N_S
Ss54	-	-	-	L	6.8 ±1.0	6.1 ±1.0	2.9 ±1.0	141786	108916	32870
Ss55	H	-	-	L	9.1 ±1.1	8.0 ±1.1	4.5 ±1.1	78096	60396	17699
Ss56	H	L	-	L	1.7 ±1.0	2.1 ±1.1	0.0 ±1.0	25108	19554	5554
Ss57	H	H	-	L	9.0 ±1.2	6.9 ±1.2	5.7 ±1.2	23853	18303	5550
Ss58	L	-	-	L	0.3 ±1.0	0.7 ±1.0	-0.2 ±1.0	63690	48519	15171
Ss59	L	L	-	L	-0.9 ±1.0	-0.4 ±1.0	-1.2 ±1.0	15290	11753	3536
Ss60	L	H	-	L	0.8 ±1.0	1.6 ±1.0	-0.7 ±1.0	21799	16481	5317
Ss61	-	L	-	L	0.8 ±1.0	1.3 ±1.0	-0.7 ±1.0	40398	31308	9090
Ss62	-	H	-	L	6.7 ±1.1	5.9 ±1.1	3.5 ±1.1	45652	34785	10867
Ss63	-	-	H	L	4.6 ±1.0	4.1 ±1.1	2.5 ±1.1	84584	63324	19852
Ss64	H	-	H	L	6.5 ±1.1	5.4 ±1.1	3.9 ±1.1	45735	34610	10503
Ss65	H	L	H	L	0.7 ±1.0	1.8 ±1.0	-0.8 ±1.0	14749	11229	3306
Ss66	H	H	H	L	5.6 ±1.2	3.5 ±1.1	3.8 ±1.2	13880	10398	3242
Ss67	L	-	H	L	0.3 ±1.0	0.3 ±1.0	0.0 ±1.0	38871	28739	9349
Ss68	L	L	H	L	-0.1 ±1.0	0.6 ±1.0	0.1 ±1.0	9234	6882	2179
Ss69	L	H	H	L	0.2 ±1.0	1.1 ±1.0	0.3 ±1.0	13278	9767	3257
Ss70	-	L	H	L	0.6 ±1.0	1.7 ±1.0	-0.7 ±1.0	23973	18115	5490
Ss71	-	H	H	L	3.9 ±1.1	3.3 ±1.1	2.8 ±1.1	27144	20160	6523
Ss72	-	-	L	L	5.0 ±1.0	4.6 ±1.1	1.5 ±1.0	57202	45592	13018
Ss73	H	-	L	L	6.5 ±1.1	6.1 ±1.1	2.3 ±1.0	32361	25785	7196
Ss74	H	L	L	L	1.8 ±1.1	1.2 ±1.0	1.1 ±1.0	10358	8325	2248
Ss75	H	H	L	L	7.4 ±1.2	6.7 ±1.2	4.5 ±1.3	9973	7904	2308
Ss76	L	-	L	L	0.0 ±1.0	0.5 ±1.0	-0.4 ±1.0	24819	19779	5822
Ss77	L	L	L	L	-1.4 ±0.9	-1.3 ±1.0	-1.9 ±0.9	6056	4871	1356
Ss78	L	H	L	L	1.0 ±1.1	1.0 ±1.0	-1.4 ±0.9	8521	6713	2059
Ss79	-	L	L	L	0.5 ±1.0	0.0 ±1.0	-0.2 ±1.0	16424	13193	3599
Ss80	-	H	L	L	6.0 ±1.1	5.3 ±1.2	2.1 ±1.1	18508	14625	4344

Cuadro 6.2: Esta tabla muestra los resultados de la medición de ζ para todos los tipos de vacíos, así como los vacíos de tipo R y S por separado para submuestras definidas como todas las combinaciones posibles de Spin alto y bajo, Masa, y Σ_5 siempre que la velocidad radial, V_{rad} , sea baja. Los radios interior y exterior de la cáscara considerada son 0.9 y $1.4R_v$ respectivamente. En general, este conjunto de submuestras con velocidades radiales bajas muestra señales de alineación más grandes.

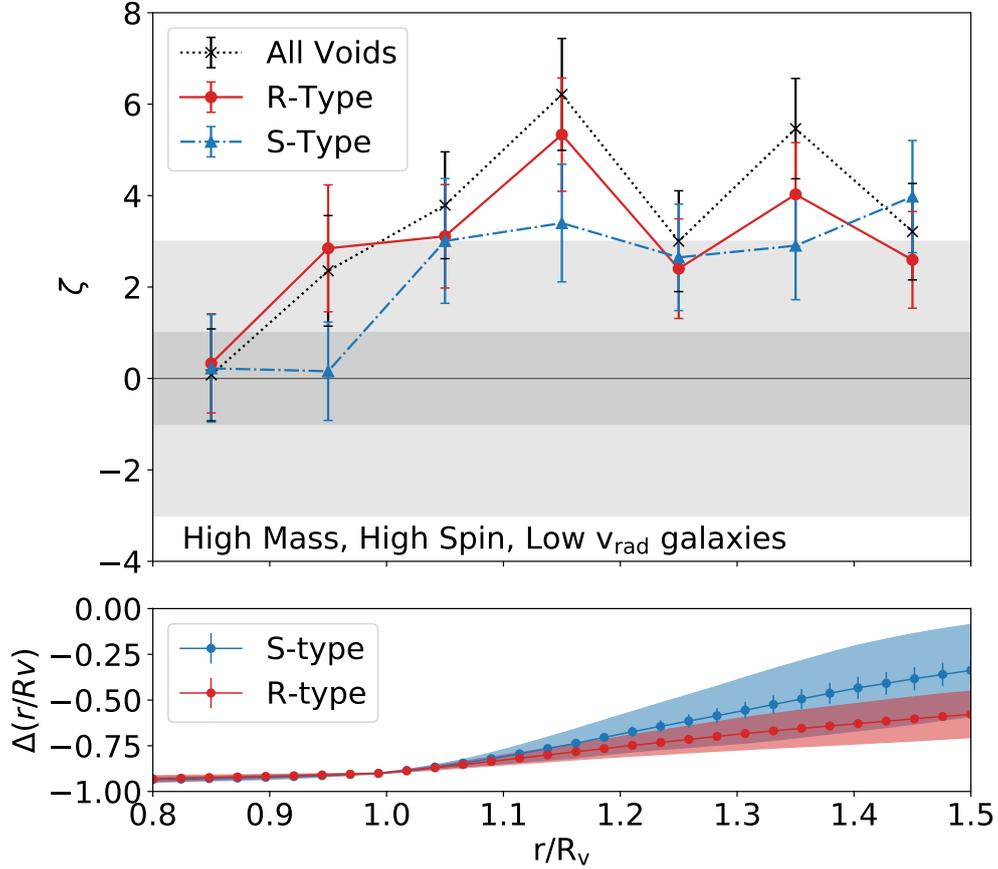


Figura 6.5: Las tablas 6.1 y 6.2 indican que las poblaciones más alineadas son galaxias masivas con alto espín y baja velocidad radial, particularmente en los vacíos de tipo R. En el panel superior de esta figura representamos la señal de alineación ζ en función de la distancia normalizada al centro del vacío. La línea negra punteada representa la señal del apilamiento de todos los vacíos, mientras que las líneas roja y azul punteadas representan la de los vacíos tipo R y S respectivamente. Encontramos un pico de señal de alineación perpendicular de más de 5σ en el intervalo de distancia centrada en $1.15R_v$ alrededor de los vacíos tipo R. El panel inferior muestra que se trata de una región muy poco densa. En este panel trazamos el contraste de densidad acumulativo en la misma escala de distancia, y es fácilmente observable que, para los vacíos tipo R, el contraste de densidad en esta escala es $\Delta(1.15R_v) \simeq -0.75$. Los puntos son las medias de $\Delta(r)$, mientras que las barras de error y las regiones sombreadas representan los errores de la media y la desviación estándar de los datos, respectivamente.

una baja velocidad de expansión respecto a los centros de los vacíos. Destacamos que esta muestra de galaxias grandes con mayor rotación son las más fiables desde el punto de vista dinámico de las alineaciones de espín.

Descubrimos que las alineaciones de los espines dependen en gran medida de la magnitud de la velocidad de expansión con respecto al centro del vacío. El hecho de que los espines más alineados sean los de galaxias con una velocidad más baja de expansión centrada en el vacío sugiere que las galaxias pueden ganar un espín alineado a medida que pierden impulso lineal en su expansión lejos del centro del vacío. En este escenario en el que las velocidades peculiares de las galaxias tienen una fuerte contribución de la expansión global del vacío, la dirección del centro del vacío es privilegiada para los encuentros de galaxias y los procesos de acreción. El hecho de que la correlación más fuerte esté relacionada con la expansión centrada en los vacíos da una pista de que las desviaciones de la dinámica global de los vacíos es un ingrediente clave para entender el origen de las alineaciones.

Por otro lado, la falta de dependencia de los resultados de la alineación con Σ_5 muestra que la densidad local de las galaxias juega un papel menor en la evolución de los vectores de espín. La inclusión de la clasificación de los vacíos proporciona más pistas sobre el origen del efecto. Nuestro análisis muestra que los vacíos de tipo R son los que exhiben los mayores efectos de alineación de espín. Esto es un indicio de que es la dinámica del vacío y su interacción con las galaxias en evolución, más que el entorno del vacío, lo que genera las orientaciones de espín sistemáticas.

En general, los estudios anteriores sobre alineaciones de espín se han relacionado con filamentos u otras estructuras sobredensas o entornos locales. Aquí detectamos la alineación en regiones subdensas, como se muestra en la Sec. 6.2.1. Nuestro hallazgo de una orientación preferentemente perpendicular es consistente con el trabajo observacional de Trujillo et al. (2006), que posteriormente fue rebatido por trabajos similares como Slosar y White (2009) y Varela et al. (2012), señalando a una muestra estadísticamente pequeña como la principal razón de la discrepancia; sin embargo, este defecto no está presente en nuestro trabajo. Además, nuestros hallazgos son consistentes con las predicciones de la Teoría de Torque Tidal (TTT) y estudios observacionales de Lee y Pen (2000, 2002) y Lee et al. (2007), donde, en este último, se calcula el campo tensorial de marea y se encuentra una alineación preferencial para los espines con su eje principal intermedio, que se encuentra dentro de las láminas (que serían un *proxy* para nuestras

superficies de vacíos). De todas formas, la señal de alineamiento que encontramos es particularmente fuerte en galaxias que se alejan de la dinámica global del vacío (ya que baja velocidad radial parece correlacionar con alineamiento), lo cual podría deberse a interacciones entre galaxias y por lo tanto se trataría de un escenario fuera del alcance de la TTT. Por otro lado, no encontramos una alineación paralela con las láminas para las galaxias masivas como la encontrada por Codis et al. (2018), Kraljic et al. (2019), muy probablemente debido a las densidades de entorno tan diferentes en las que residen estas galaxias. Además, observamos que, al tener en cuenta las galaxias masivas, encontramos fuertes señales de alineación solo para aquellas con alto espín, i.e. no encontramos ninguna señal de alineación significativa para las galaxias de alta masa y bajo espín (véanse las submuestras Ss3 y Ss6 en la Tabla 6.1). Aunque este efecto podría deberse a la determinación más precisa del eje de espín en el caso de las galaxias de alto espín, también podría indicar una diferencia importante entre las galaxias con alto y bajo cociente espín–masa. Sin embargo, observamos que una comparación directa entre algunos de estos trabajos y el presente es difícil de evaluar, ya que no hemos realizado un cálculo del campo del tensor de marea, y la dirección vacío–céntrica sólo puede tomarse como una aproximación estadística a la dirección del eje principal mayor.

Capítulo 7

Estructura con escalamiento jerárquico en vacíos cósmicos

Volviendo al escenario más amplio de la estructura en gran escala, este capítulo aborda una de las motivaciones generales de esta Tesis que es la de buscar estadísticos alternativos a los tradicionales para caracterizar propiedades y estructuras de galaxias. Aquí, además de hablar de las estructuras conocidas como vacíos cósmicos, volvemos a hablar de los vacíos estrictos: volúmenes completamente desprovistos de galaxias. Por ello, retomamos la nomenclatura de “vacío cósmico” para referirnos a la componente de la Red Cósmica, y llamamos simplemente “vacíos” a los vacíos estrictos del recuento en celdas.

En la Sec. 3.3 vimos que la función de probabilidad de vacío (VPF) es la probabilidad P_0 de no encontrar objetos en una celda de un dado volumen, y es un caso particular del enfoque de recuento en celdas (*counts-in-cells*, CiC) para la caracterización de la distribución de galaxias. Vimos que a partir de un campo inicial de fluctuaciones gaussianas de densidad, y con soluciones que asumen “autosimilaridad” y “agrupamiento estable”, se obtiene una distribución de galaxias tal que las funciones de correlación de altos órdenes, ξ_p con $p > 2$, pueden escribirse en términos de la tradicional función de correlación bipuntual, ξ , por medio de coeficientes adimensionales de escala: $S_p \simeq \xi_p/\xi^{p-1}$.

Cuando introducimos este Ansatz, llamado “escalamiento jerárquico” o “escala jerárquica”, en la formulación de la VPF obtenemos una relación entre la probabilidad de encontrar cero galaxias a un radio R y todos los órdenes superiores de correlación. Esta formulación nos permite escribir la VPF en función de una variable de escala $\bar{N}\bar{\xi}$, que puede entenderse como el número de vecinos en

exceso al caso aleatorio para cualquier galaxia. Se han utilizado diversos modelos fenomenológicos de agrupamiento con orígenes en distintas áreas de la física, donde la forma de la VPF es conocida, para comparar con la agrupación de galaxias, siendo los más exitosos el modelo binomial negativo y el termodinámico. En general se ha encontrado que el modelo binomial negativo ajusta bien la distribución de galaxias en el espacio de *redshift*, mientras que el termodinámico o lognormal ajusta, con menos fidelidad, la distribución de la materia oscura. Un acuerdo con uno u otro modelo jerárquico puede ser indicativo de escalamiento jerárquico en la muestra de estudio. Finalmente, para eliminar la dependencia de la VPF con la densidad numérica de una o más muestras de estudio, se introduce la VPF reducida, o RVPF, como $\chi(\bar{N}\bar{\xi}) = -\ln(P_0)/\bar{N}$ que, como puede observarse, toma valores $0 < \chi < 1$ donde el valor máximo corresponde a una distribución poissoniana.

El escalamiento jerárquico requiere que los coeficientes no dependan de la escala, y como vimos en la Sec. 3.3.5 esto es, en el mejor de los casos, sólo aproximadamente cierto. Sin embargo hay numerosos trabajos que verifican la relación de escala jerárquica de la distribución de galaxias en el espacio de *redshift*, donde el modelo jerárquico binomial negativo parece ser un buen ajuste a los datos (ver Sec. 3.3.4).

En este capítulo reportamos y comparamos resultados del estudio de la función de probabilidad de vacío reducida dentro y fuera de vacíos cósmicos identificados en TNG300-1, tanto en el espacio real como en el espacio de *redshift*.

7.1. Metodología

A continuación describimos la metodología general para obtener estadísticas de vacíos en muestras de galaxias. La investigación de la VPF (y la RVPF) requiere la medición de tres magnitudes: \bar{N} , $\bar{\xi}$, y P_0 , todas ellas funciones del radio de la esfera, R . Cada una de estas cantidades se determina directamente mediante un enfoque de recuento en celdas. Simplemente se coloca un gran número N_{esf} de esferas aleatorias en los datos, en nuestro caso la simulación TNG300-1, y se cuenta el número de galaxias que contiene cada esfera. Esto se repite para muchos radios de esfera. \bar{N} es el número medio de galaxias en una esfera,

$$\bar{N} = \frac{1}{N_{\text{esf}}} \sum_{i=1}^{N_{\text{esf}}} N_i \quad (7.1)$$

donde N_i es el número de galaxias en la i -ésima esfera. P_0 es el número de esferas sin galaxias N_0 dividido por el número total de esferas N_{esf} ,

$$P_0 = \frac{N_0}{N_{\text{esf}}} \quad (7.2)$$

y $\bar{\xi}$ es la varianza en el número de galaxias por esfera:

$$\bar{\xi} = \frac{\overline{(N - \bar{N})^2} - \bar{N}}{\bar{N}^2}. \quad (7.3)$$

En el límite de un gran número de puntos aleatorios, se sabe que el enfoque de recuento en celdas para determinar $\bar{\xi}$ es matemáticamente equivalente a los métodos más convencionales (Szapudi y Szalay, 1998). Conroy et al. (2005) confirmaron este resultado de forma independiente comparando la estimación de $\bar{\xi}$ con el método CiC con la función de correlación promediada por volumen obtenida mediante el popular estimador de Landy-Szalay, encontrando que los dos enfoques son equivalentes.

Este trabajo consiste en calcular la RVPF dentro y fuera de vacíos cósmicos identificados en la simulación TNG300-1. La identificación de vacíos cósmicos es la misma que la presentada en la Sec. 6.2. La finalidad es comparar estos estadísticos y estudiar cuál, si lo hubiere, es el efecto del entorno de los vacíos cósmicos en el agrupamiento de galaxias.

7.1.1. Pruebas de estabilidad de los estadísticos

Para estudiar la sensibilidad de estas cantidades al número total de esferas utilizadas, N_{esf} , hemos calculado P_0 , \bar{N} , y $\bar{\xi}$ como función de la cantidad de esferas para una muestra de puntos de distribución aleatoria en una caja de lado unidad. Los resultados se muestran en el conjunto de paneles izquierdo de la Fig. 7.1. Los paneles derechos muestran el mismo estudio pero en la TNG300-1. Las líneas negras a trazos representan algunas cantidades esperadas como referencia. La cantidad de puntos n de la muestra aleatoria es 10^7 , del mismo orden que la cantidad de galaxias en la simulación TNG300-1, y el radio de las esferas en ambos casos es apenas menor que la separación media entre partículas, calculada

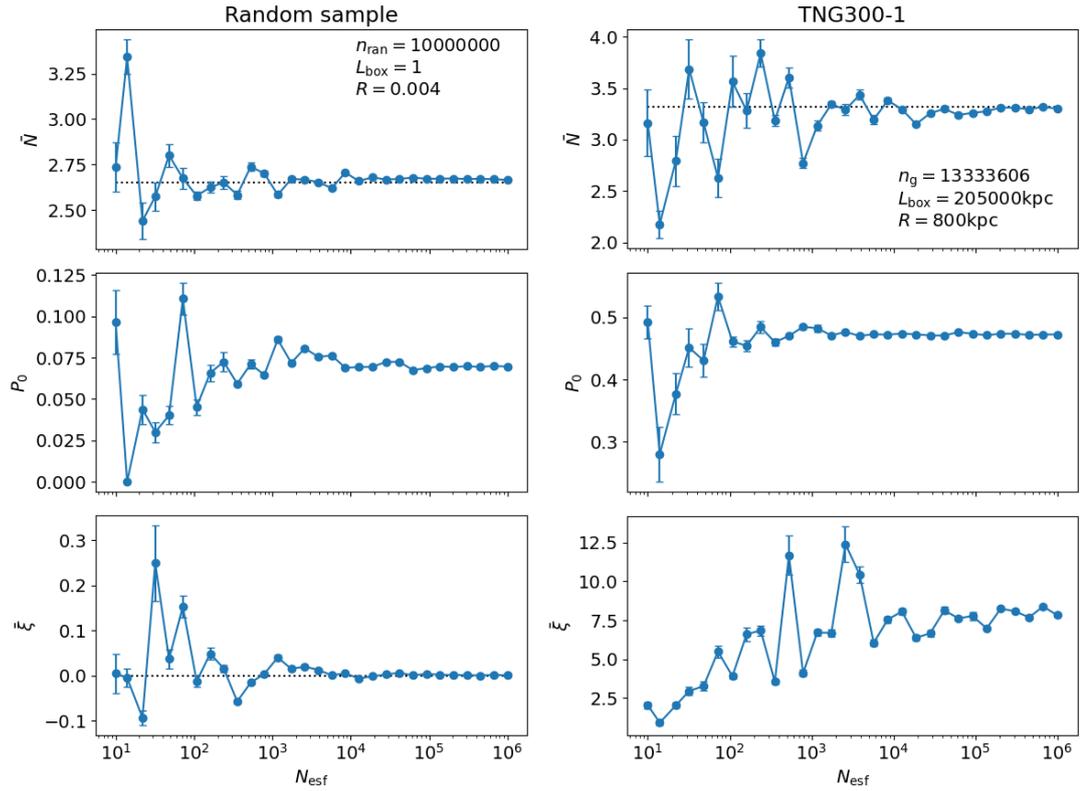


Figura 7.1: Prueba de estabilidad de los estadísticos P_0 , \bar{N} y $\bar{\xi}$, calculados mediante recuentos en celdas, respecto al número de esferas de muestreo. En el conjunto de paneles derechos tenemos los estadísticos calculados en la simulación, mientras que en los izquierdos están calculados en una muestra control de puntos posicionados aleatoriamente, donde la cantidad de puntos es del orden de la cantidad de galaxias en la simulación. La estimación de los estadísticos se estabilizan para una cantidad de esferas de muestreo $N_{\text{esf}} > 10^4$. Elegimos emplear $N_{\text{esf}} = 10^5$.

como $r_{\text{msep}} = (V_{\text{box}}/N)^{1/3}$, donde V_{box} es el volumen completo de la caja o *box* y N es el número de partículas o galaxias en ella. Con esta elección del radio de muestreo nos aseguramos un valor de $P_0 > 0$.

Como es esperable, para N_{esf} pequeños, estas cantidades son inestables, pero para $N_{\text{esf}} > 10^4$, las cantidades convergen bien. Para asegurarnos una estima confiable de estas cantidades, utilizamos $N_{\text{esf}} = 10^5$.

Los errores de \bar{N} y $\bar{\xi}$ se estimaron mediante un remuestreo Jackknife del volumen total. Sin embargo, para calcular las incertezas de P_0 y χ se utilizaron expresiones derivadas analíticamente (Colombi et al., 1994, Hamilton, 1985, Maurogordato y Lachieze-Rey, 1987):

$$\Delta P_0 = \sqrt{\frac{P_0(1 - P_0)}{N_{\text{esf}}}}, \quad (7.4)$$

y

$$\left(\frac{\Delta\chi}{\chi}\right) \simeq \left| \frac{\Delta P_0}{P_0|\ln P_0|} - \frac{\Delta\bar{N}}{\bar{N}} \right|, \quad (7.5)$$

ya que al calcular la incerteza de $\chi = -\ln P_0/\bar{N}$, el denominador y numerador no son independientes sino que están casi exactamente anticorrelacionados (ver Colombi et al., 1994, Fry y Colombi, 2013).

Para calcular los estadísticos dentro de los vacíos cósmicos identificados en la TNG300-1, muestreamos uniformemente el volumen esférico de estos. El método de identificación y las propiedades generales de los vacíos cósmicos están presentados en la Sec. 6.2. Estos vacíos tienen un rango de radios de $7 - 11h^{-1}\text{Mpc}$. Teniendo en cuenta esto, determinamos que el radio de los volúmenes de muestreo dentro de los vacíos cósmicos sea menor a la mitad del radio mínimo de estos. El rango de radios de las esferas de muestreo dentro de los vacíos es de $0.25 - 2.5h^{-1}\text{Mpc}$. De esta forma, el gran número de esferas de muestreo no impide medir de forma independiente los estadísticos como \bar{N} o $\bar{\xi}$. El rango de radios de las esferas de muestreo para el cálculo de la VPF en toda la caja de la simulación es de $0.04 - 4h^{-1}\text{Mpc}$.

7.1.2. Periodicidad y simulación del espacio de *redshift*

Como la simulación es periódica en su volumen, sus bordes están tan muestreados como el centro de la caja, independientemente del radio de las esferas

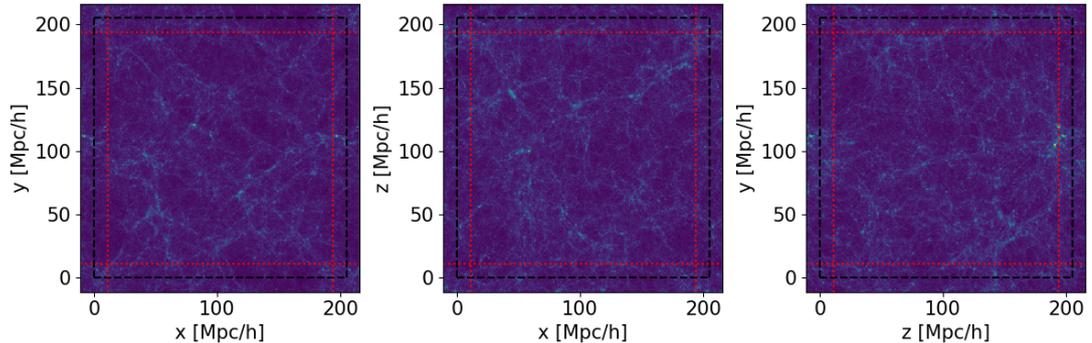


Figura 7.2: Esta figura muestra la replicación periódica de los bordes de la caja de la simulación TNG300-1 en proyección en los 3 ejes. El borde negro a trazos representa la proyección del volumen original, mientras que las líneas rojas punteadas representan los márgenes que están siendo replicados en los bordes opuestos. En la práctica, el ancho del borde a replicar es igual al radio de las esferas de muestreo.

de muestreo. En la práctica nos aseguramos de esto extendiendo la simulación periódicamente hasta los límites $[L_{\text{box}} - R, L_{\text{box}} + R]$, donde R es el radio de las esferas de muestreo. En la Fig. 7.2 mostramos proyecciones en distintos ejes donde delineamos con trazos negros los límites originales de la simulación, y con líneas punteadas rojas los bordes que son replicados en los límites opuestos; puede verificarse visualmente la periodicidad cruzando la línea de trazos negros.

Además, para simular el espacio de *redshift* en la TNG300-1 utilizamos la aproximación plano-paralelo, que consiste en asumir que el observador se encuentra infinitamente lejos y por lo tanto podemos ajustar las posiciones de las galaxias a lo largo de un sólo eje. Hemos calculado los estadísticos presentados más adelante modificando los tres ejes de la caja y hemos comprobado que los resultados son indistinguibles entre sí. Empleamos el parámetro $H_0 = 0.06774(km/s)/kpc$ de la TNG300-1 y hacemos la conversión $x_f = x_i + v_x/H_0$, donde x y v_x son la coordenada y velocidad en el eje x respectivamente, y aplicamos condiciones periódicas si corresponde (ver Fig. 7.3).

7.2. Resultados

En esta sección repasamos los resultados de calcular la RVPF (Ec. 3.14) en TNG300-1 en el espacio real, espacio de *redshift*, y dentro y fuera de los vacíos cósmicos.

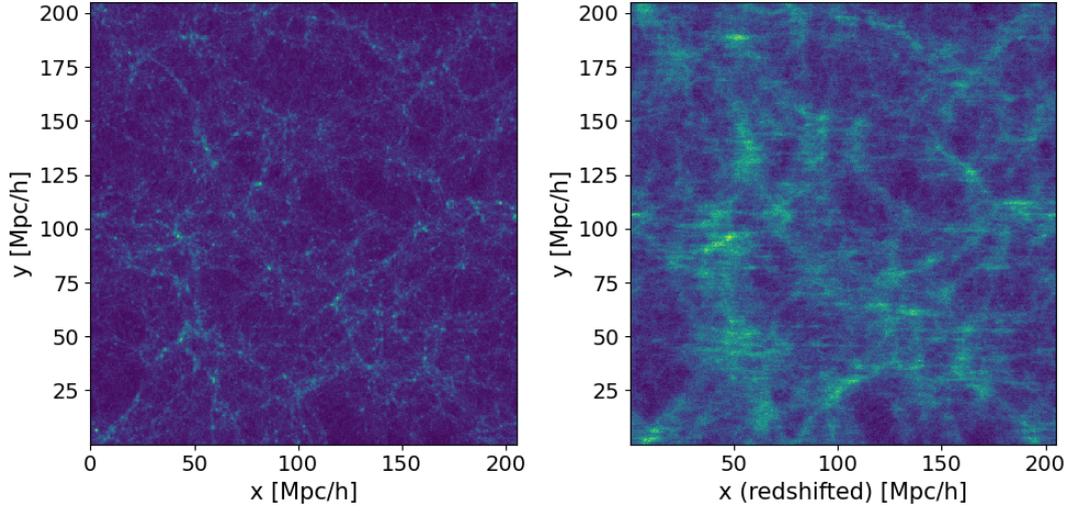


Figura 7.3: En esta figura mostramos el resultado de simular un espacio de *redshift* en TNG300-1, empleando la aproximación plano-paralelo con un observador situado en $x = -\infty$ y alterando el eje x de la simulación con la velocidad peculiar de las galaxias y el parámetro H_0 . Los resultados son estadísticamente idénticos afectando cualquiera de los 3 ejes.

Comenzamos comparando los resultados en todo el *box* de la simulación en el espacio real y en el espacio de *redshift*, ya que en este último es donde está reportado un aparente acuerdo con el escalamiento jerárquico (ver Sec. 3.3). Previo a mostrar la RVPF en función de su variable de escala, $\chi(\bar{N}\bar{\xi})$, es ilustrativo analizar cómo se relacionan las cantidades intermedias, \bar{N} , $\bar{\xi}$ y P_0 , entre sí y cómo varían con la escala.

7.2.1. RVPF en el espacio real vs. en el espacio de *redshift*

Para familiarizarnos con la naturaleza de los estadísticos de recuentos en celdas, o estadísticos CiC, calculamos \bar{N} , $\bar{\xi}$, P_0 , y la mencionada variable de escala $\bar{N}\bar{\xi}$ en todo el *box* como función de los radios de muestreo R que, como mencionamos más arriba, tienen un rango de $0.04 - 4h^{-1}\text{Mpc}$. La Fig. 7.4 muestra los resultados, donde las curvas azules representan los estadísticos en el espacio real y las curvas rojas en el espacio de *redshift*.

Como es de esperar, la media de la cantidad de galaxias en las esferas de muestreo aumenta con el radio y son idénticas en ambos espacios. La varianza de este número, i.e. la correlación promedio $\bar{\xi}$, y la VPF, o P_0 , naturalmente decrecen con la escala.

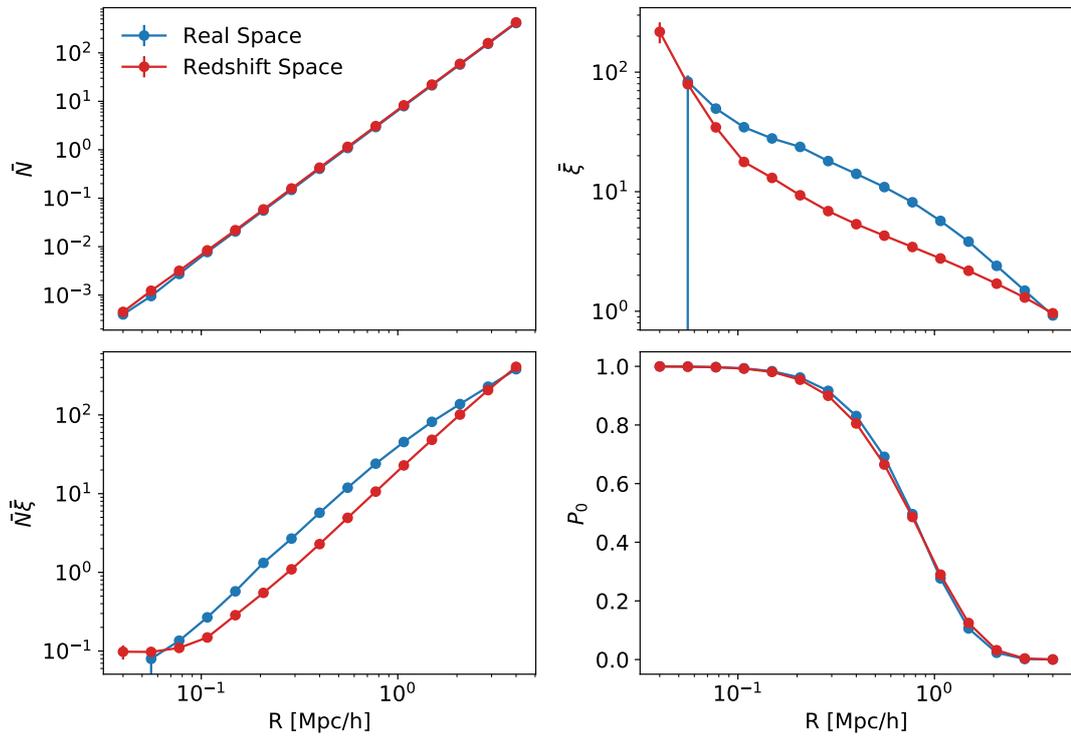


Figura 7.4: Comparación de la variación de los estadísticos CiC con la escala en los espacios real y de *redshift*. En las pequeñas escalas que estamos considerando, el efecto predominante en el espacio de *redshift* es el de los dedos de dios, que difumina sobredensidades disminuyendo la amplitud de la correlación. Observamos que $\bar{N}\xi$ es una variable creciente con la escala.

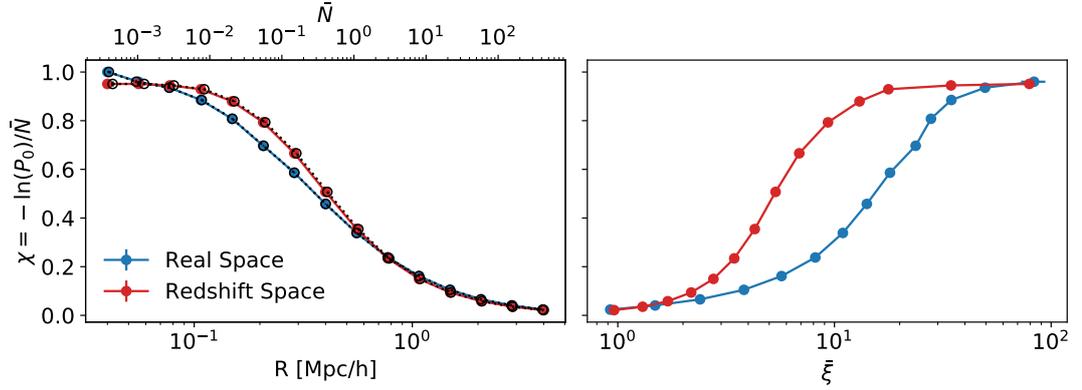


Figura 7.5: Relación de la RVPF, χ , con el radio R (panel izquierdo) y los constituyentes de la variable de escala \bar{N} y $\bar{\xi}$ (panel derecho). Las funciones $\chi(R)$ y $\chi(\bar{N})$ están graficadas en el mismo panel ya que sus comportamientos son casi idénticos, con ésta última dibujada en líneas negras a trazos y círculos vacíos. Naturalmente, la RVPF decrece con R , pero aumenta con la correlación $\bar{\xi}$.

Como vimos en la Sec. 3.2, hay dos efectos presentes en el espacio de *redshift*: el efecto Kaiser a gran escala que resulta en un mayor tamaño aparente de los vacíos, y el efecto dedos de dios en escalas pequeñas que difuminan estructuras sobredensas. En este análisis estamos limitados a escalas pequeñas, por lo que el efecto predominante será el de los dedos de dios. Esto se ve en la menor amplitud de la correlación en el espacio de *redshift*, que resulta en valores levemente menores también de P_0 en escalas $\lesssim 1h^{-1}\text{Mpc}$; los efectos en el espacio de *redshift* difuminan grandes aglomeraciones y por lo tanto disminuye la correlación de galaxias así como la probabilidad de encontrar vacíos.

El comportamiento de la variable $\bar{N}\bar{\xi}$ con la escala difiere para los dos espacios dadas las diferencias en $\bar{\xi}$, pero es útil ver que, en cualquier caso, si bien $\bar{N}\bar{\xi}$ puede tomar distintos valores en ambos espacios para un mismo radio de muestreo, ésta es una cantidad que aumenta con R .

Luego, analizamos el comportamiento de la RVPF, χ , con respecto al radio y a las componentes de la variable de escala: \bar{N} y $\bar{\xi}$ (Fig. 7.5). Un efecto interesante se ve cuando calculamos $\chi(R)$. Al normalizar por la densidad media, vemos que la probabilidad de encontrar esferas vacías de hecho *aumenta* en el espacio de *redshift* con respecto al espacio real para un dado radio R . La variación de la RVPF con la densidad media \bar{N} es muy similar a su comportamiento respecto a R , por lo que decidimos graficarlo en el mismo panel con círculos vacíos y líneas

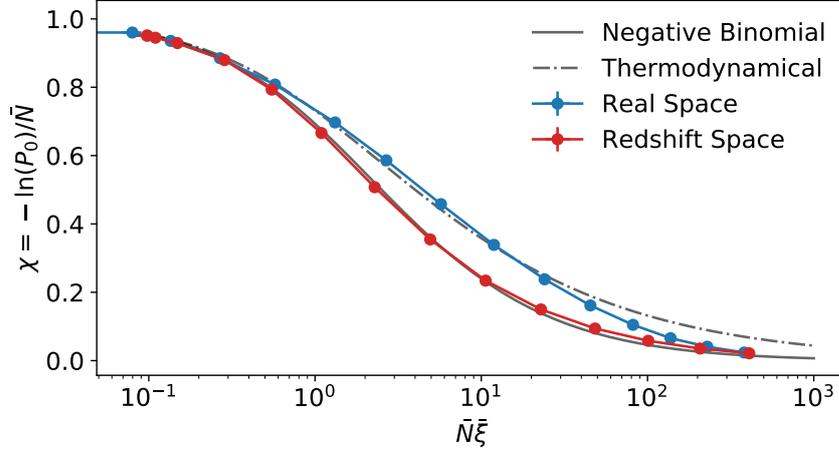


Figura 7.6: RVPF de las galaxias en TNG300-1. Vemos el buen ajuste del modelo binomial para la distribución de galaxias en el espacio de *redshift* como ha sido reportado previamente. Sin embargo se ve también un aparente acuerdo entre galaxias en el espacio real y el modelo termodinámico para $\bar{N}\bar{\xi} \lesssim 30$ (este acuerdo no se sostiene rigurosamente con diluciones, ver Fig. 7.7). Ambas curvas coinciden para valores grandes de $\bar{N}\bar{\xi}$.

negras punteadas.

Por otro lado, χ aumenta con $\bar{\xi}$, lo cual es esperable dado que cuando una muestra está más agrupada es más fácil encontrar vacíos. Sin embargo, para un dado valor de la correlación de galaxias, la RVPF es mayor en el espacio de *redshift* que en el real, y consecuentemente tiende a su valor máximo, $\chi = 1$, para valores menores de correlación que en el espacio real.

Finalmente, graficamos la RVPF, χ , en función de la variable de escala, $\bar{N}\bar{\xi}$, en la Fig. 7.6. Vemos que en el espacio de *redshift* las galaxias siguen un modelo jerárquico binomial negativo, como se ha reportado en trabajos previos. Además, encontramos un sorprendentemente buen acuerdo entre la RVPF de las galaxias en el espacio real y el modelo termodinámico en valores de $\bar{N}\bar{\xi} \lesssim 30$. Para valores mayores de la variable de escala, las galaxias en el espacio real parecen volver a tender al modelo binomial, donde coincide con el espacio de *redshift*.

Para probar la robustez de estos resultados, seguimos el análisis de Croton et al. (2004b) y diluimos la muestra de forma aleatoria para verificar si la relación de escala se mantiene. Estos resultados se muestran en la Fig. 7.7, donde mostramos el espacio real en el conjunto izquierdo de paneles y el espacio de *redshift* en el derecho. En ambos casos los paneles superiores exhiben la muestra completa

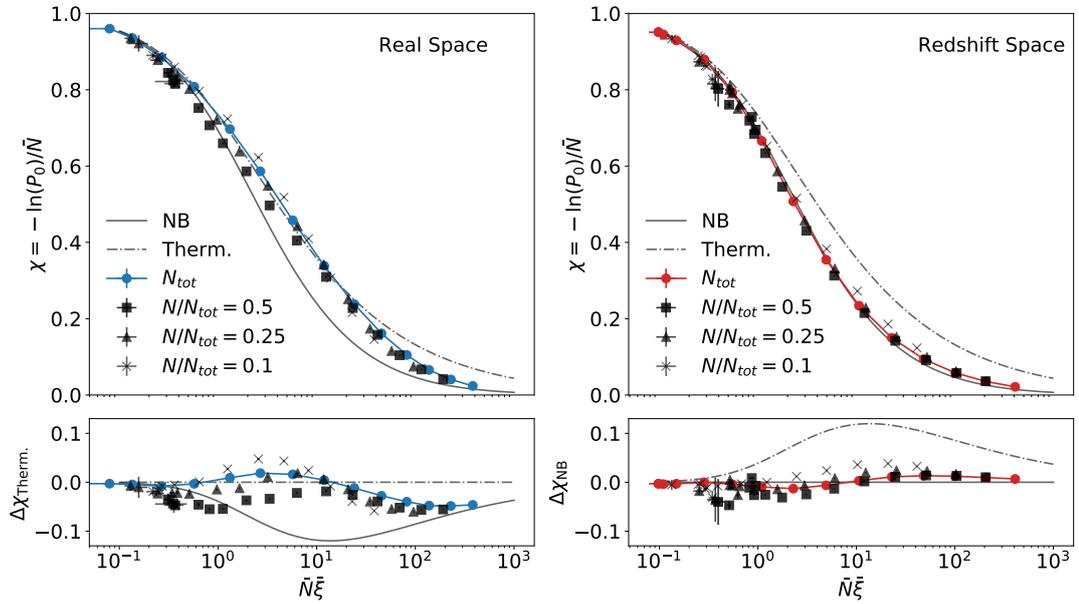


Figura 7.7: RVPF de las galaxias en TNG300-1 en el espacio de *redshift* (paneles derechos) y real (paneles izquierdos) con diluciones aleatorias del 50, 25 y 10%. Los paneles inferiores muestran las restas entre la RVPF de las muestras con la RVPF del modelo termodinámico ($\Delta\chi_{Therm.}$, panel inferior izquierdo) y con la del modelo binomial negativo ($\Delta\chi_{NB}$, panel inferior derecho). El escalamiento jerárquico se sostiene mejor en el espacio de *redshift* que en el real, donde se ve en los paneles inferiores que exhibe una menor dispersión entre las diluciones.

con N_{tot} en color y diluciones del 50, 25, y 10 % con cuadrados, triángulos y cruces negras con transparencia incremental respectivamente. Los paneles inferiores muestran las cantidades

$$\Delta\chi_{\text{Therm.}} \equiv \chi - \chi_{\text{Therm.}}$$

y

$$\Delta\chi_{\text{NB}} \equiv \chi - \chi_{\text{NB}},$$

correspondientes a las diferencias entre la RVPF de las muestras y las RVPF de los modelos termodinámico y binomial negativo respectivamente. Se puede observar que la dispersión entre las diluciones es mayor en el espacio real que en el espacio de *redshift*; el escalamiento jerárquico en el espacio real no se sostiene tan rigurosamente como en el espacio de *redshift*.

La Fig. 7.7 puede compararse con la Fig. 3.10 extraída del trabajo de [Vogeley et al. \(1994\)](#) donde compara la RVPF en el espacio de *redshift* y en el espacio real en simulaciones de distintos modelos cosmológicos. Los comportamientos son similares para la simulación que denota como CDM1, con $\Omega_m = 1$ y $h = 5$ (ver Sec. 3.3.4), aunque estos parámetros son, por supuesto, distintos a los de TNG300-1 que describimos en la Sec. 6.1.

7.2.2. RVPF dentro de vacíos cósmicos en el espacio real vs. en el espacio de *redshift*

En esta sección analizamos los estadísticos calculados en la sección anterior, pero dentro de vacíos cósmicos identificados en TNG300-1 con el motivo de analizar si hay alguna diferencia en el agrupamiento de las galaxias dentro de los mismos. El rango de escalas para estudiar estos entornos es un poco más acotado, con radios de $0.25 - 2.5h^{-1}\text{Mpc}$, por lo que el rango en \bar{N} y $\bar{\xi}$ es menor aquí también.

La Fig. 7.8 muestra los estadísticos CiC en función del radio R. Al igual que en la subsección anterior, las curvas azules y rojas representan el espacio real y de *redshift* respectivamente, pero empleamos círculos vacíos para diferenciar y remarcar el hecho de que aquí estamos estudiando regiones subdensas.

Notamos un comportamiento invertido respecto a la subsección anterior en la comparación espacio de *redshift* y real. Vemos ahora que en estos entornos subdensos la correlación $\bar{\xi}$ (panel superior derecho) y la variable de escala $\bar{N}\bar{\xi}$ (inferior izquierdo) son mayores en el espacio de *redshift* que en el real en todo el

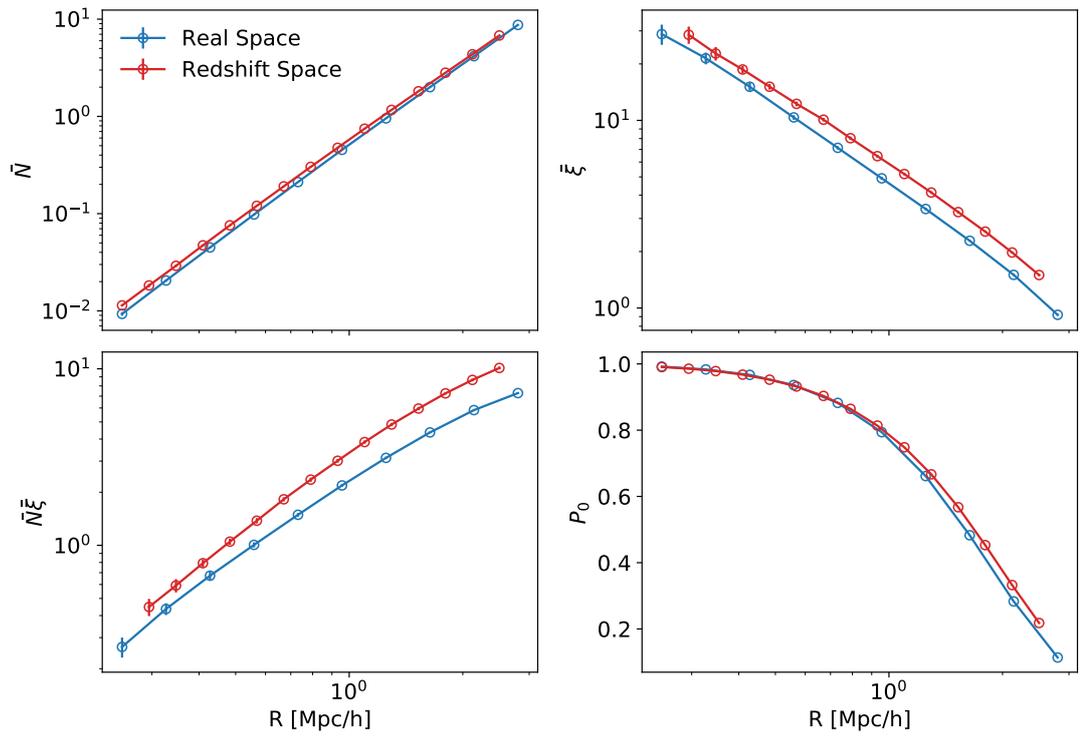


Figura 7.8: Comparación de la variación de los estadísticos CiC con la escala en los espacios real y de *redshift*, dentro de vacíos cósmicos. Los estadísticos \bar{N} , $\bar{\xi}$ y $\bar{N}\bar{\xi}$ son menores en el espacio real que en el espacio de *redshift* en toda la escala considerada. La P_0 es marginalmente superior en el espacio de *redshift* en escalas $R \gtrsim 1h^{-1}\text{Mpc}$.

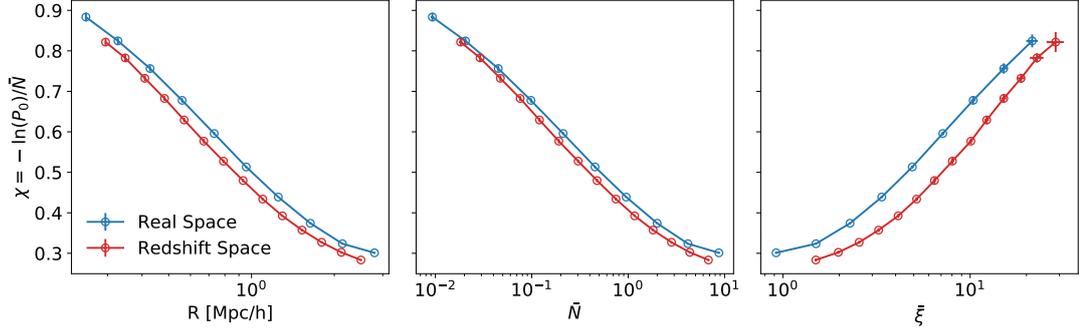


Figura 7.9: χ vs estadísticos CiC en vacíos cósmicos

rango de radios considerado. Las VPFs (panel inferior derecho) son indistinguibles hasta $R \lesssim 1h^{-1}\text{Mpc}$, y para escalas mayores la P_0 en el espacio de *redshift* es marginalmente superior.

Al calcular la variación de χ con R y \bar{N} notamos que tienen comportamientos similares pero no tan indistinguibles como en la subsección anterior, por lo que en la Fig. 7.9 los graficamos en paneles separados (izquierdo y central, respectivamente). El panel derecho, $\chi(\bar{\xi})$ refleja lo mismo que el panel superior derecho de la Fig. 7.8, donde vemos que $\bar{\xi}$ tiene un comportamiento similar en ambos espacios pero con distinta amplitud.

Finalmente, al calcular la RVPF en función de la variable de escala $\bar{N}\bar{\xi}$ (Fig. 7.10), vemos que tanto en el espacio real como en el de *redshift*, las galaxias dentro de los vacíos cósmicos parecen agruparse según el modelo binomial negativo. Aún más sorprendente es el hecho de que este ajuste parece ser mejor en el caso del espacio real.

Reproduciendo el análisis de la subsección anterior, diluimos aleatoriamente las muestras de galaxias para probar la robustez de estos resultados. Podemos ver en la Fig. 7.11 que si bien ambos espacios presentan poca dispersión, esta es aún menor en el espacio real. Aunque el modelo binomial parece sobreestimar levemente los datos, este buen acuerdo es inesperado dado que trabajos previos indican que cualquier escalamiento jerárquico encontrado en el espacio de *redshift* se rompe en el espacio real.

7.2.3. Comparaciones finales

Para salir de la comparación espacio real vs. *redshift* en la que hemos estructurado la información hasta ahora, buscamos mostrar más directamente las

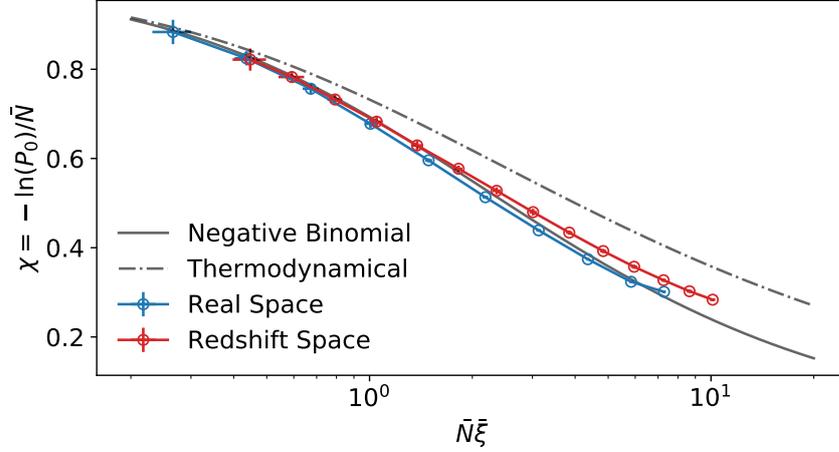


Figura 7.10: RVPF dentro de vacíos cósmicos en el espacio de *redshift* y real. Ambas muestras se ajustan al modelo binomial negativo. Contrario a la RVPF en toda la caja, las galaxias dentro de los vacíos cósmicos en el espacio real parecen ajustarse bien al modelo binomial en todo el rango de escalas considerado, mientras que en el espacio de *redshift* se desvía de este modelo para valores $\bar{N}\bar{\xi} \gtrsim 2$.

diferencias en las RVPFs dentro y fuera de los vacíos cósmicos.

La Fig. 7.12 muestra las curvas dentro y fuera de los vacíos cósmicos en cada panel, donde, al igual que antes, ubicamos el espacio real en los paneles izquierdos y el espacio de *redshift* en los derechos. Recordamos que los círculos vacíos representan los estadísticos dentro de los vacíos cósmicos.

En el espacio de *redshift* parece haber un desvío de la escala jerárquica dentro de los vacíos cósmicos para valores $\bar{N}\bar{\xi} \gtrsim 1$, pero tanto dentro como fuera de los mismos, las galaxias parecen seguir a grandes rasgos el modelo binomial negativo. Esto es en líneas generales lo que se ha reportado en la literatura hasta el momento.

Sorprendentemente, para el espacio real es muy distinto calcular la RVPF dentro y fuera de los vacíos cósmicos. Vimos que fuera de ellos, las galaxias parecen seguir el modelo termodinámico, aunque este escalamiento se rompe al diluir la muestra. Sin embargo, dentro de los vacíos cósmicos parece haber un muy buen acuerdo con el modelo binomial negativo. Este buen ajuste se mantiene aún con las diluciones.

Cabe aclarar que al diluir aleatoriamente la muestra completa, los vacíos cósmicos, por definición, son las estructuras menos afectadas. Si bien esto for-

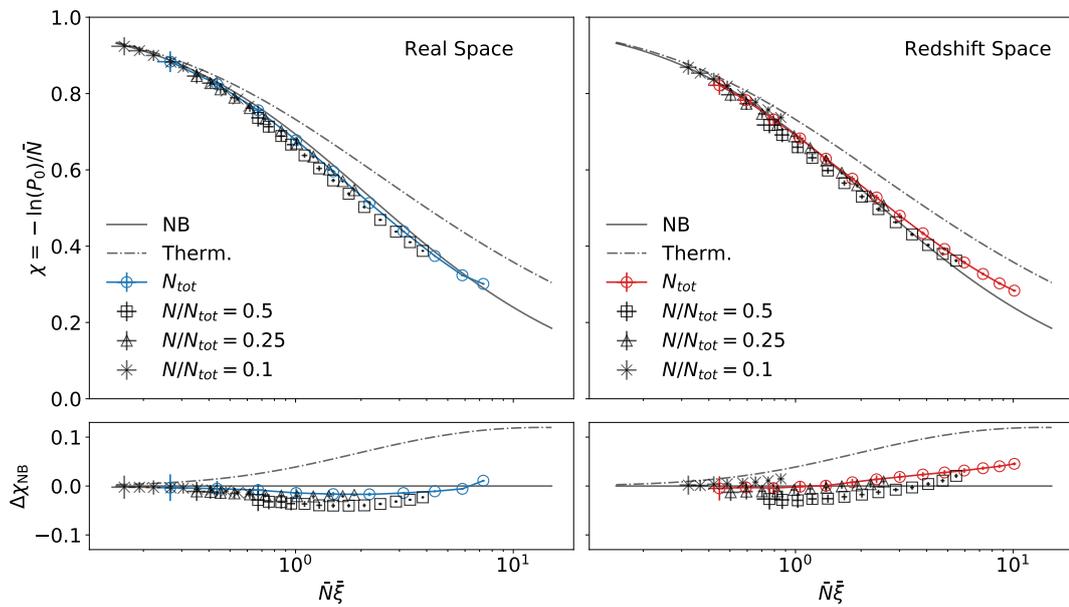


Figura 7.11: RVPF de las galaxias dentro de vacíos cósmicos en TNG300-1 en el espacio de *redshift* (paneles derechos) y real (paneles izquierdos) con diluciones aleatorias del 50, 25 y 10%. Los paneles inferiores muestran las restas entre la RVPF de las muestras con la RVPF del modelo termodinámico ($\Delta\chi_{Therm.}$, panel inferior izquierdo) y con la del modelo binomial negativo ($\Delta\chi_{NB}$, panel inferior derecho). Contrario a resultados previos, la poca dispersión en las RVPFs diluidas en el panel inferior izquierdo indica una posible detección de escalamiento jerárquico de las galaxias dentro de vacíos cósmicos en el espacio real.

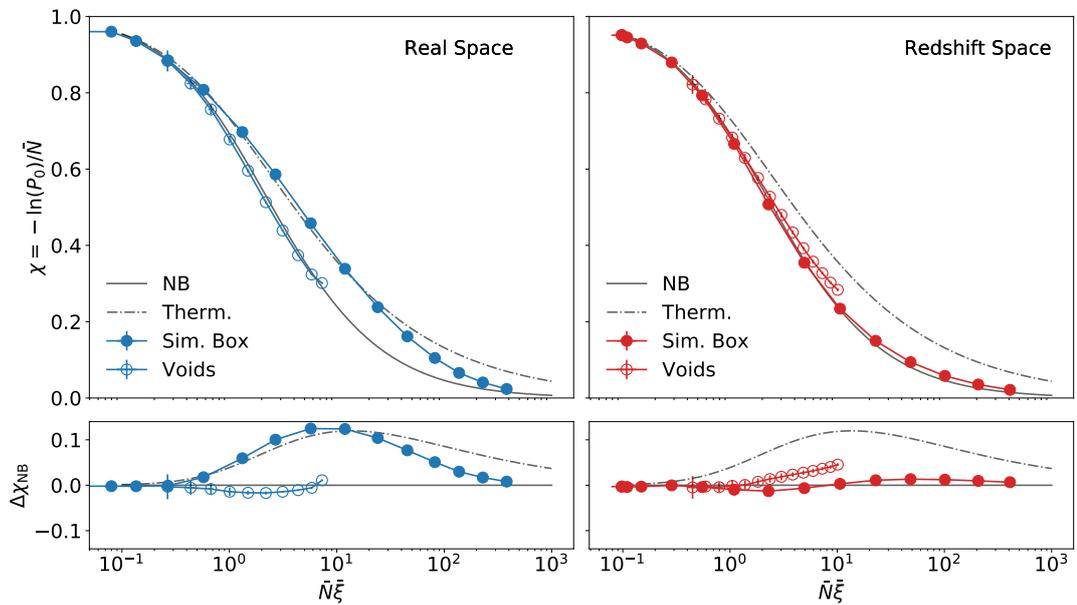


Figura 7.12: RVPF dentro y fuera de vacíos cósmicos en el espacio real y de *redshift*. Los símbolos vacíos representan la RVPF calculada dentro de vacíos cósmicos mientras que los sólidos representan el cálculo en toda la simulación. Vemos que calcular la RVPF dentro y fuera de vacíos cósmicos en el espacio de *redshift* no es muy distinto, pero en el espacio real la diferencia es notable. La RVPF dentro de vacíos cósmicos en el espacio real se ajusta al modelo binomial negativo, al igual que la RVPF en el espacio de *redshift*.

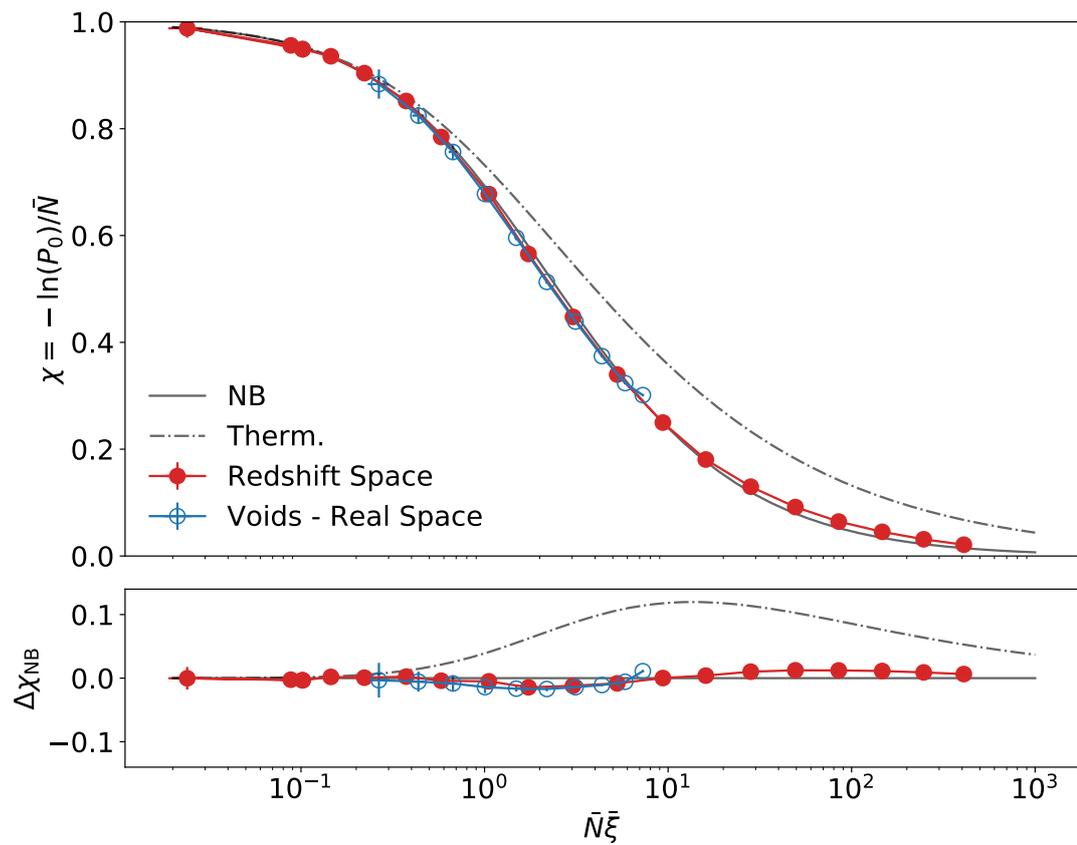


Figura 7.13: RVPF en el espacio de *redshift* y en vacíos cósmicos en el espacio real. Recuperamos el comportamiento de escalamiento jerárquico reportado en relevamientos en el espacio de *redshift* pero dentro de los vacíos cósmicos en el espacio real

ma parte de la razón por la cual los estadísticos se ven menos afectados dentro de los vacíos cósmicos, no debería ser el causante principal de este buen comportamiento. Por un lado, la RVPF por definición está desafectada de la densidad media en cada esfera de muestreo, lo cual en principio la hace un buen estadístico para usar en distintos entornos. Por otro lado, la dilución máxima es del 10 % del total de la muestra, es decir que es comparable con la densidad interna de un vacío cósmico (ver criterios y algoritmos de identificación en la Sec. 6.2); aún con esta dilución, la RVPF fuera de los vacíos cósmicos está lejos de ajustarse al modelo binomial negativo que parece caracterizar tanto la distribución dentro de ellos en el espacio real, como dentro y fuera en el espacio de *redshift*.

A partir de estas ideas, mostramos finalmente en la Fig. 7.13 la RVPF en el espacio de *redshift* con la RVPF en el espacio real dentro de los vacíos cósmicos. El sorprendente acuerdo entre ambas indica que dentro de los vacíos cósmicos hay indicios de escalamiento jerárquico que no se encuentran en otros entornos del Universo.

7.3. Conclusiones

Hemos calculado estadísticos de vacío en la simulación TNG300-1, dentro y fuera de vacíos, y en el espacio real y de *redshift*. Comparamos la RVPF de las galaxias en estas muestras con dos modelos populares de agrupamiento: el modelo binomial negativo y el termodinámico.

Encontramos que en el espacio de *redshift* la agrupación de galaxias exhibe un comportamiento acorde al escalamiento jerárquico correspondiente al modelo binomial negativo en todo el rango de escalas considerado. Este buen acuerdo con algún modelo jerárquico se rompe cuando calculamos la RVPF en el espacio real, lo cual está en acuerdo con estudios previos (Lahav et al., 1993, Vogeley et al., 1994). En el espacio real encontramos que los datos ocupan un locus aproximado al modelo termodinámico en valores de $\bar{N}\bar{\xi} \lesssim 20$, aunque con diferencias de hasta $\chi - \chi_{\text{Term.}} \simeq \pm 0.05$ en las diluciones más extremas, para luego tomar valores menores y más aproximados al modelo binomial negativo.

Al repetir el cálculo dentro de vacíos cósmicos, encontramos que aún en el espacio real, donde esperamos que el escalamiento jerárquico se deshaga, hay un excelente acuerdo entre la RVPF de las galaxias en vacíos cósmicos y el modelo binomial negativo, con diferencias máximas de $\chi - \chi_{\text{BN}} \simeq \pm 0.01$ para la muestra

total y -0.03 para una dilución del 50 %. La RVPF dentro de vacíos cósmicos en el espacio real es sorprendentemente parecida a aquella calculada para toda la simulación en el espacio de *redshift*; i.e. el comportamiento consistente con escalamiento jerárquico en el espacio de *redshift* es recuperado dentro de vacíos cósmicos en espacio real.

Es interesante contrastar este resultado con el argumento de Croton et al. (2006) que postula que el movimiento peculiar de las galaxias no debe ser el principal agente en generar un escalamiento jerárquico consistente con el binomial negativo. Además, este resultado también puede entenderse como una consecuencia de que los contrastes de densidades son menores dentro de los vacíos, y por lo tanto las soluciones de “autosimilaridad” y “agrupamiento estable” son válidas en estos regímenes; en los vacíos cósmicos podemos encontrar escalamiento jerárquico que se deshace en ambientes más densos y pueden encontrarse estructuras que conservan de forma más pristina las condiciones iniciales del Universo.

Capítulo 8

Conclusiones

Con esta Tesis nos propusimos estudiar algunas propiedades importantes de las galaxias en relación a vacíos cósmicos, siempre desde un punto de vista estadístico. Esta labor nos llevó desde el estudio de propiedades dinámicas como el alineamiento de espines galácticos con respecto a la dirección radial de vacíos cósmicos y su dependencia con propiedades astrofísicas de las galaxias, hasta estudiar la naturaleza de la distribución de galaxias dentro de vacíos cósmicos, pasando por la presentación de nuevos métodos estadísticos pertinentes al área.

Para presentar las conclusiones, vamos a revisar las preguntas postuladas en el inicio:

- ¿Puede mejorarse el estudio de la distribución de las galaxias en el espacio y su orientación con respecto a estructuras circundantes?
- ¿Existen métodos alternativos a estos? ¿Son igualmente válidos? ¿Qué información aportan y qué nos dicen acerca de las estructuras en vacíos cósmicos?
- ¿La orientación de las galaxias es aleatoria? ¿Cuál es su relación con estructuras en gran escala como los vacíos cósmicos?

Primero, abordamos la práctica más general del estudio del agrupamiento de galaxias en el Universo e intentamos mejorarlo. En el Cap. 4 estudiamos el impacto sobre la precisión de los estimadores de la función de correlación bipuntual al sustituir el catálogo aleatorio utilizado en la estimación por distribuciones de puntos tipo *glass*. Demostramos que es posible generar distribuciones de partículas de este tipo aplicando iterativamente la reconstrucción de Zeldovich a un conjunto

de puntos inicialmente aleatorio. El pequeño costo de cómputo adicional asociado a la construcción de dicha muestra se ve compensado por la mejora significativa en la precisión de las estimaciones de la función de correlación.

Proporcionamos una versión modificada del estimador LS, adaptada al uso de distribuciones de partículas tipo *glass* (Ec. 4.9). Este estimador hace uso de dos catálogos *glass* independientes para evitar problemas debidos a la correlación de los puntos dentro de una misma muestra para escalas que se acercan a la separación media entre partículas. Nuestros resultados indican que al usar distribuciones tipo *glass* no añadimos ningún sesgo significativo respecto a los resultados obtenidos utilizando muestras Poisson del mismo tamaño. En escalas mayores a la separación media entre partículas el uso de catálogos tipo *glass* conduce a una reducción significativa de la varianza de los multipolos $\xi_\ell(s)$ con respecto a los resultados del estimador LS estándar con el mismo número de puntos. El tamaño del catálogo *glass* necesario para lograr una dada precisión en la función de correlación es significativamente menor que cuando se utilizan muestras aleatorias. Como el costo computacional de estos estimadores es proporcional al número total de pares, el menor tamaño del catálogo *glass* puede representar una reducción significativa del tiempo computacional total. Por ejemplo, si extrapolamos el comportamiento de ley de potencia de la varianza σ_{ξ_ℓ} mostrado en la Fig. 4.7 encontramos que la misma varianza conseguida por el estimador LS utilizando muestras aleatorias de $\alpha \equiv N_r/N_d = 50$ como se hizo en los análisis finales del BOSS (Alam et al., 2017, Sánchez et al., 2017) se puede obtener utilizando catálogos *glass* con $\alpha = 5.4$. En los dos ejemplos anteriores, la reducción del tiempo de cómputo que se obtiene utilizando catálogos *glass* en lugar de catálogos aleatorios asciende a $t_R/t_G \simeq 86$ y $t_R/t_G \simeq 145$, respectivamente, lo que representa una gran reducción de los recursos computacionales necesarios para el análisis. Esta mejora puede resultar particularmente beneficiosa para las mediciones de las BAO luego de la reconstrucción, ya que requieren recuentos de pares en catálogos aleatorios con y sin la aplicación del campo de desplazamiento.

Por último, la Sec. 4.4 contiene versiones de otros estimadores comúnmente utilizados de la $\xi(\mathbf{s})$ y de N -puntos adaptados al uso de catálogos *glass*. Los estimadores que emplean muestras *glass* podrían resultar cada vez más útiles en el cálculo de estadísticas de alto orden, donde el costo computacional de contar N -tuplas de puntos podría verse drásticamente reducido sin comprometer la varianza. Aunque en el cálculo del espectro de potencia el tamaño del catálogo

aleatorio no es el factor principal para determinar el costo computacional, la estadística del espacio de Fourier también podría beneficiarse de la menor varianza de los catálogos *glass* tanto en la estimación de $P(k)$ como en la función ventana del relevamiento. En los próximos años, relevamientos como DESI y Euclid proporcionarán catálogos con decenas de millones de objetos en grandes volúmenes. El análisis de estos catálogos representará un reto para las técnicas de análisis tradicionales, y el uso de catálogos *glass* podría ayudar a reducir los requisitos computacionales de los análisis de agrupación en estos estudios y sus catálogos simulados asociados, manteniendo la alta precisión que exigen.

En el Cap. 5, siguiendo la línea de estadísticos nuevos, presentamos dos métodos para detectar y cuantificar las señales de alineación en galaxias y comprobar su robustez estadística. El primer método se basa en la utilización de un parámetro, $\eta \equiv N(\beta > 1)/N(\beta < 1)$, donde $\beta \equiv \tan(\lambda)$ y λ es el ángulo agudo entre las direcciones en estudio. Este parámetro η se puede interpretar como la fracción relativa de componentes vectoriales en el plano perpendicular a una dirección de referencia y derivamos los primeros dos momentos de la distribución de este estimador, pudiendo así evaluar de forma fiable su significación estadística sin depender de simulaciones Monte Carlo. El parámetro η tiene un valor de expectación y una varianza dadas por $\eta_0 \simeq 2,4142$ y $\text{Var}(\eta) \simeq 28,1421/N$.

El segundo método que presentamos realiza un ajuste sobre los residuos de la ECDF de los datos con respecto a la ECDF esperada para una distribución uniforme. El ajuste utiliza un pequeño conjunto de funciones armónicas ortogonales y no se basa en ningún esquema de discretización. La amplitud del ajuste, i.e., la amplitud de la señal de alineación, puede ser descrita por el primer parámetro impar del ajuste, a_1 .

El segundo método de análisis de alineación que presentamos produce coeficientes OLS de un ajuste de los residuos de la ECDF de los cosenos de los datos con respecto a una muestra aleatoria. El primer coeficiente impar, a_1 , de la expansión armónica de la función residual es suficiente para caracterizar la amplitud de la señal de alineación. Los valores positivos de a_1 indican una alineación perpendicular, valores nulos son consistentes con alineamientos aleatorios, y valores negativos indican una alineación paralela con respecto a la dirección privilegiada.

Comparamos estos métodos con la media de la distribución de los cosenos y empleando simulaciones de Monte Carlo de la misma, que es la forma más común de cuantificar alineaciones. Esta comparación se realizó generando puntos alea-

toriamente en la superficie de elipsoides con distintas excentricidades, simulando así una población de vectores con tres niveles diferentes de alineación (desde la no alineación, a la intermedia, a la muy alineada), y analizando qué tan fidedigna es la recuperación de la señal de alineación tanto con los parámetros nuevos como con el método más clásico del cálculo de la distribución de cosenos.

Encontramos que los estadísticos propuestos permiten la detección de señales de alineación con una mayor significación. Para una desviación de aproximadamente $0,25\sigma$ de una distribución isotrópica de cosenos, obtenemos una significación de $10-$ y 12σ para el coeficiente a_1 y el parámetro η respectivamente.

Por último, evaluamos los efectos de las incertidumbres en la medición de las componentes vectoriales paralelas y perpendiculares introduciendo ruido gaussiano en las mismas simulando errores de observación. Encontramos una relación lineal entre los parámetros de alineamiento obtenidos con los datos sintéticos, y los parámetros “observados” que tienen en cuenta el error de observación simulado. Esto puede utilizarse para estimar el tamaño necesario de una muestra para lograr la detección de una señal pequeña en los datos. Con estas herramientas, los próximos grandes relevamientos y simulaciones pueden proporcionar nuevos conocimientos a partir de pequeñas señales de alineamientos que no se detectan en los relevamientos actuales con un número menor de galaxias.

En el Cap. 6 utilizamos las herramientas presentadas en el Cap. 5 y analizamos las orientaciones de los espines galácticos en entornos subdensos en la simulación Illustris TNG300-1. Los resultados muestran una fuerte evidencia de que las grandes galaxias en vacíos cósmicos exhiben un exceso de espines perpendiculares a la dirección radial céntrica del vacío.

Encontramos la mayor señal de alineación ($\sim 9\sigma$) para galaxias masivas, con gran rotación, que residen en entornos de vacío y, en particular, con una baja velocidad de expansión respecto a los centros de vacío. El hecho de que la correlación más fuerte esté relacionada con la expansión centrada en los vacíos da una pista de que las desviaciones de la dinámica global de los vacíos es un ingrediente clave para entender el origen de las alineaciones. Este resultado también sugiere que las galaxias pueden ganar un espín alineado a medida que pierden impulso lineal en su expansión lejos del centro del vacío. Además, implica que la dirección del centro del vacío podría ser privilegiada para los encuentros de galaxias y los procesos de acreción. Esto sería interesante estudiar en futuros trabajos.

Por otro lado, los resultados muestran que la densidad local de las galaxias

juega un papel menor en la evolución de los vectores de espín. Incluimos en nuestro análisis la dependencia de la señal con el tipo de vacío cósmico y concluimos que los vacíos de tipo R son los que exhiben los mayores efectos de alineación de espín. Esto es un indicio de que es la dinámica del vacío y su interacción con las galaxias en evolución, más que el entorno del vacío, lo que genera las orientaciones de espín sistemáticas.

Finalmente, en el Cap. 7 abordamos el cálculo de la función de probabilidad de vacío reducida (RVPF) en la simulación TNG300-1, dentro y fuera de vacíos, y en el espacio real y de *redshift*, y comparamos estas muestras con dos modelos populares de agrupamiento: el modelo binomial negativo y el termodinámico. Este estadístico es mucho menos usual que los otros estadísticos de agrupación de galaxias nombrados en esta tesis, e.g. la función de correlación bipuntual. La literatura al respecto parece indicar que la RVPF por sí sola puede ser insuficiente para describir de forma completa la distribución de galaxias, sin embargo es un estadístico complementario que forma parte de un enfoque con un gran potencial para extraer información de relevamientos y simulaciones como es el de recuento de celdas.

En nuestro análisis, acorde a resultados previos, encontramos que en el espacio de *redshift* la agrupación de galaxias exhibe un comportamiento acorde al escalamiento jerárquico correspondiente al modelo binomial negativo en todo el rango de escalas considerado. Sin embargo, el escalamiento no se mantiene en el espacio real: encontramos que los datos se ajustan aproximadamente al modelo termodinámico siempre que la variable de escala $\bar{N}\bar{\xi}$ no tome valores muy grandes, donde los datos parecen tender al modelo binomial negativo. Notamos, además, que el acuerdo con el modelo termodinámico en el espacio real no se mantiene al diluir la muestra, mientras que el escalamiento jerárquico en el espacio de *redshift*, ajustado por el modelo binomial, sí se mantiene aún con diluciones.

Sorprendentemente, al realizar el cálculo de la RVPF dentro de vacíos cósmicos, encontramos que en el espacio real, donde esperamos que el escalamiento jerárquico se deshaga, hay un excelente acuerdo entre la RVPF de las galaxias en vacíos cósmicos y el modelo binomial negativo, aún cuando diluimos la muestra. La RVPF dentro de vacíos cósmicos en espacio real es notablemente similar a la calculada en toda la simulación en espacio de *redshift*; i.e. el escalamiento jerárquico detectado en espacio de *redshift* es recuperado dentro de vacíos cósmicos en espacio real.

En los vacíos cósmicos podemos detectar la presencia de estructura con escalamiento jerárquico. Este escalamiento se deshace en ambientes más densos, por lo que los resultados indican que en los vacíos cósmicos pueden encontrarse estructuras que conservan de forma más pristina condiciones más tempranas del Universo.

Bibliografía

- Kevork Abazajian, Jennifer K. Adelman-McCarthy, Marcel A. Agueros, Sahar S. Allam, Kurt Anderson, Scott F. Anderson, James Annis, et al. The Second Data Release of the Sloan Digital Sky Survey. *AJ*, 128(1):502–512, 2004. doi: 10.1086/421365.
- Shadab Alam, Metin Ata, Stephen Bailey, Florian Beutler, Dmitry Bizyaev, Jonathan A Blazek, Adam S Bolton, et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *MNRAS*, 470(3):2617–2652, 2017. doi: 10.1093/mnras/stx721.
- M. A. Aragon-Calvo, R. van de Weygaert, P. A. Araya-Melo, E. Platen, y A. S. Szalay. Unfolding the hierarchy of voids. *MNRAS*, 404(1):L89–L93, 2010. doi: 10.1111/j.1745-3933.2010.00841.x.
- Miguel A. Aragon-Calvo. The Hierarchical Nature of the Spin Alignment of Dark Matter Haloes in Filaments. *MNRAS: Letters*, 440(1):L46, 2013. doi: 10.1093/mnrasl/slu009. URL <http://arxiv.org/abs/1303.1590><http://dx.doi.org/10.1093/mnrasl/slu009>.
- Miguel A. Aragón-Calvo, Rien van de Weygaert, Bernard J. T. Jones, y J. M. van der Hulst. Spin Alignment of Dark Matter Halos in Filaments and Walls. *ApJ*, 655(1):L5–L8, 2007. ISSN 0004-637X. doi:10.1086/511633.
- B. Aryal, W. Saurer, B. Aryal, y W. Saurer. Comments on the expected isotropic distribution curves in galaxy orientation studies. *A&A*, 364:L97–L100, 2000. ISSN 0004-6361. URL <https://ui.adsabs.harvard.edu/abs/2000A&A...364L..97A/abstract>.
- Horace W. Babcock. The rotation of the Andromeda Nebula. *Lick Observatory Bulletin*, 498:41–51, 1939. doi:10.5479/ADS/bib/1939LicOB.19.41B.

- Steven A. Balbus y John F. Hawley. Instability, turbulence, and enhanced transport in accretion disks. *Reviews of Modern Physics*, 70(1):1–53, 1998. doi: 10.1103/RevModPhys.70.1.
- R. Balian, R. Schaeffer, R. Balian, y R. Schaeffer. Scale-invariant matter distribution in the universe. *A&A*, 220:1–29, 1989. ISSN 0004-6361. URL <https://ui.adsabs.harvard.edu/abs/1989A&A...220...1B/abstract>.
- Michael Balzer, Thomas Schlömer, y Oliver Deussen. Capacity-constrained point distributions. *ACM Transactions on Graphics*, 28(3):1, 2009. ISSN 07300301. doi:10.1145/1531326.1531392. URL <http://portal.acm.org/citation.cfm?doid=1576246.1531392>.
- C. M. Baugh, D. J. Croton, E. Gaztañaga, P. Norberg, M. Colless, I. K. Baldry, J. Bland-Hawthorn, et al. The 2dF galaxy redshift survey: Hierarchical galaxy clustering. *MNRAS*, 351(2):L44–L48, 2004. ISSN 00358711. doi: 10.1111/j.1365-2966.2004.07962.x. URL <https://ui.adsabs.harvard.edu/abs/2004MNRAS.351L..44B/abstract>.
- C. M. Baugh y G. Efstathiou. The three-dimensional power spectrum measured from the APM galaxy survey - I. Use of the angular correlation function. *MNRAS*, 265:145–156, 1993. doi:10.1093/mnras/265.1.145.
- Julian E Bautista, Mariana Vargas-Magaña, Kyle S Dawson, Will J Percival, Jonathan Brinkmann, Joel Brownstein, Benjamin Camacho, et al. The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations at Redshift of 0.72 with the DR14 Luminous Red Galaxy Sample. *ApJ*, 863(1):110, 2018. doi:10.3847/1538-4357/aacea5.
- Eric Jones Baxter y Eduardo Rozo. A Maximum Likelihood Approach to Estimating Correlation Functions. *ApJ*, 779(1), 2013. ISSN 15384357. doi: 10.1088/0004-637X/779/1/62. URL <https://ui.adsabs.harvard.edu/abs/2013ApJ...779...62B/abstract><http://arxiv.org/abs/1305.4613><http://dx.doi.org/10.1088/0004-637X/779/1/62>.
- C. Benoist, A. Cappi, L. N. da Costa, S. Maurogordato, F. R. Bouchet, y R. Schaeffer. Biasing and High-Order Statistics from the Southern-Sky Redshift Survey. *ApJ*, 514(2):563–578, 1999. ISSN 0004-637X. doi:10.1086/306955. URL <https://ui.adsabs.harvard.edu/abs/1999ApJ...514..563B/abstract>.

- F. Bernardeau, S. Colombi, E. Gaztañaga, y R. Scoccimarro. Large-scale structure of the Universe and cosmological perturbation theory. *Physics Report*, 367(1-3):1–248, 2002. ISSN 03701573. doi:10.1016/S0370-1573(02)00135-7. URL <https://ui.adsabs.harvard.edu/abs/2002PhR...367...1B/abstractwww.elsevier.com/locate/physrep>.
- E. Bertschinger. The self-similar evolution of holes in an Einstein-de Sitter universe. *ApJS*, 58:1–37, 1985. doi:10.1086/191027.
- Juan Betancort-Rijo, Santiago G. Patiri, Francisco Prada, y Antonio Enea Romano. The statistics of voids as a tool to constrain cosmological parameters: σ_8 and Γ . *MNRAS*, 400(4):1835–1849, 2009. ISSN 13652966. doi:10.1111/J.1365-2966.2009.15567.X. URL <http://arxiv.org/abs/0901.1609><http://dx.doi.org/10.1111/j.1365-2966.2009.15567.x>.
- Chris Blake, Alexandra Pope, Douglas Scott, y Bahram Mobasher. On the cross-correlation of sub-mm sources and optically selected galaxies. *MNRAS*, 368(2):732–740, 2006. doi:10.1111/j.1365-2966.2006.10158.x.
- G. R. Blumenthal, S. M. Faber, R. Flores, y J. R. Primack. Contraction of Dark Matter Galactic Halos Due to Baryonic Infall. *ApJ*, 301:27, 1986. doi:10.1086/163867.
- J. Richard Bond, Lev Kofman, y Dmitry Pogosyan. How filaments of galaxies are woven into the cosmic web. *Nature*, 380(6575):603–606, 1996. doi:10.1038/380603A0.
- S. A. Bonometto, S. Borgani, S. Ghigna, A. Klypin, y J. R. Primack. Non-linear clustering in the cold plus hot dark matter model. *MNRAS*, 273(1):101–121, 1993. ISSN 0035-8711. doi:10.1093/mnras/273.1.101. URL <http://arxiv.org/abs/astro-ph/9311018><http://dx.doi.org/10.1093/mnras/273.1.101>.
- Francois R. Bouchet. Moments of the Counts Distribution in the 1.2 Jy IRAS Redshift Survey. *ApJ*, 417:36, 1993. doi:10.1086/173289. URL <http://arxiv.org/abs/astro-ph/9305018><http://dx.doi.org/10.1086/173289>.
- Francois R. Bouchet, Michael A. Strauss, Marc Davis, Karl B. Fisher, Amos Yahil, y John P. Huchra. Moments of the Counts Distribution in the 1.2 Jansky IRAS Galaxy Redshift Survey. *ApJ*, 417:36, 1993. doi:10.1086/173289.

- Tereasa G. Brainerd. Anisotropic Distribution of SDSS Satellite Galaxies: Planar (Not Polar) Alignment. *ApJL*, 628(2):L101–L104, 2005. doi:10.1086/432713.
- M-A. Breton y S. de la Torre. Fast analytical random pair counts calculation for realistic survey geometry. *arXiv e-prints*, arXiv:2010.02793, 2020.
- Riccardo Brunino, Ignacio Trujillo, Frazer R. Pearce, y Peter A. Thomas. The orientation of galaxy dark matter haloes around cosmic voids. *MNRAS*, 375(1):184–190, 2007. doi:10.1111/j.1365-2966.2006.11282.x.
- A. Burden, W. J. Percival, y C. Howlett. Reconstruction in Fourier space. *MNRAS*, 453(1):456–468, 2015. ISSN 0035-8711. doi:10.1093/mnras/stv1581. URL <https://academic.oup.com/mnras/article-lookup/doi/10.1093/mnras/stv1581>.
- P. Carruthers y Duong-Van Minh. The evidence for and meaning of the universality of hadronic multiplicities. *Physics Letters B*, 114(2-3):169–171, 1982. doi:10.1016/0370-2693(82)90139-3.
- P. Carruthers y C. C. Shih. Correlations and fluctuations in hadronic multiplicity distribution: The meaning of KNO scaling. *Physics Letters B*, 127(3-4):242–250, 1983. doi:10.1016/0370-2693(83)90884-5.
- L. Ceccarelli, N. D. Padilla, C. Valotto, y D. G. Lambas. Voids in the 2dFGRS and Λ CDM simulations: Spatial and dynamical properties. *MNRAS*, 373(4):1440–1450, 2006. doi:10.1111/J.1365-2966.2006.11129.X.
- L. Ceccarelli, D. Paz, M. Lares, N. Padilla, y D. Garcia Lambas. Clues on void evolution - I. Large-scale galaxy distributions around voids. *MNRAS*, 434(2):1435–1442, 2013. doi:10.1093/mnras/stt1097.
- S. Codis, R. Gavazzi, Y. Dubois, C. Pichon, K. Benabed, V. Desjacques, D. Pogosyan, et al. Intrinsic alignment of simulated galaxies in the cosmic web: Implications for weak lensing surveys. *MNRAS*, 448(4):3391–3404, 2015. ISSN 13652966. doi:10.1093/MNRAS/STV231.
- S. Codis, A. Jindal, N. E. Chisari, D. Vibert, Y. Dubois, C. Pichon, y J. Devriendt. Galaxy orientation with the cosmic web across cosmic time. *MNRAS, Volume 481, Issue 4, p.4753-4774*, 481(4):4753, 2018. ISSN 0035-

8711. doi:10.1093/MNRAS/STY2567. URL <https://ui.adsabs.harvard.edu/abs/2018MNRAS.481.4753C/abstract>.
- Alison L. Coil. The large-scale structure of the universe. *Planets, Stars and Stellar Systems: Volume 6: Extragalactic Astronomy and Cosmology*, 6:387–421, 2013. doi:10.1007/978-94-007-5609-0{_}8. URL <https://ui.adsabs.harvard.edu/abs/2013pss6.book..387C/abstract>.
- Jorg M. Colberg, Ravi K. Sheth, Antonaldo Diaferio, Liang Gao, y Naoki Yoshida. Voids in a Λ CDM universe. *MNRAS*, 360(1):216–226, 2005. doi:10.1111/j.1365-2966.2005.09064.x.
- Jörg M. Colberg, Frazer Pearce, Caroline Foster, Erwin Platen, Riccardo Brunino, Mark Neyrinck, Spyros Basilakos, et al. The aspen–amsterdam void finder comparison project. *MNRAS*, 387(2):933–944, 2008. doi:10.1111/j.1365-2966.2008.13307.x. URL <https://doi.org/10.1111%2Fj.1365-2966.2008.13307.x>.
- Shaun Cole, Will J Percival, John A Peacock, Peder Norberg, Carlton M Baugh, Carlos S Frenk, Ivan Baldry, et al. The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *MNRAS*, 362(2):505–534, 2005. doi:10.1111/j.1365-2966.2005.09318.x.
- Peter Coles y Bernard Jones. A lognormal model for the cosmological mass distribution. *MNRAS*, 248:1–13, 1991. doi:10.1093/mnras/248.1.1.
- Matthew Colless, Gavin Dalton, Steve Maddox, Will Sutherland, Peder Norberg, Shaun Cole, Joss Bland-Hawthorn, et al. The 2dF Galaxy Redshift Survey: spectra and redshifts. *MNRAS*, 328(4):1039–1063, 2001. doi:10.1046/j.1365-8711.2001.04902.x.
- Matthew Colless, Gavin Dalton, Steve Maddox, Will Sutherland, Peder Norberg, Shaun Cole, Joss Bland-Hawthorn, et al. The 2dF Galaxy Redshift Survey: Spectra and redshifts. *MNRAS*, 328(4):1039–1063, 2001. ISSN 00358711. doi:10.1046/j.1365-8711.2001.04902.x.
- S. Colombi, F. R. Bouchet, y R. Schaeffer. A Count Probability Cookbook: Spurious Effects and the Scaling Model. *ApJSupplement Series*, 96:401, 1994. doi:10.1086/192125. URL <http://arxiv.org/abs/astro-ph/9409052http://dx.doi.org/10.1086/192125>.

- Charlie Conroy, Alison L. Coil, Martin White, Jeffrey A. Newman, Renbin Yan, Michael C. Cooper, Brian F. Gerke, et al. The DEEP2 Galaxy Redshift Survey: The Evolution of Void Statistics from $z \sim 1$ to $z \sim 0$. *ApJ*, 635(2):990–1005, 2005. doi:10.1086/497682.
- Asantha Cooray y Ravi Sheth. Halo models of large scale structure. *PhysRep*, 372(1):1–129, 2002. doi:10.1016/S0370-1573(02)00276-4.
- Rupert A. C. Croft y Christopher A. Metzler. Weak-Lensing Surveys and the Intrinsic Correlation of Galaxy Ellipticities. *ApJ*, 545(2):561–571, 2000. doi:10.1086/317856.
- D. J. Croton, E. Gaztañaga, C. M. Baugh, P. Norberg, M. Colless, L. K. Baldry, J. Bland-Hawthorn, et al. The 2dF Galaxy Redshift Survey: higher-order galaxy correlation functions. *Mon. Not. R. Astron. Soc*, 352(4):1232–1244, 2004a. ISSN 00358711. doi:10.1111/j.1365-2966.2004.08017.x. URL <https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2004.08017.x>.
- Darren J. Croton, Matthew Colless, Enrique Gaztañaga, Carlton M. Baugh, Peder Norberg, I. K. Baldry, J. Bland-Hawthorn, et al. The 2dF Galaxy Redshift Survey: voids and hierarchical scaling models. *MNRAS*, 352(3):828–836, 2004b. ISSN 0035-8711. doi:10.1111/j.1365-2966.2004.07968.x. URL <https://academic.oup.com/mnras/article/352/3/828/1211985https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2004.07968.x>.
- Darren J. Croton, Peder Norberg, Enrique Gaztanaga, y Carlton M. Baugh. Statistical analysis of galaxy surveys - III: The non-linear clustering of red and blue galaxies in the 2dFGRS. *MNRAS*, 379(4):1562–1570, 2006. doi:10.1111/j.1365-2966.2007.12035.x. URL <http://arxiv.org/abs/astro-ph/0611313http://dx.doi.org/10.1111/j.1365-2966.2007.12035.x>.
- Antonio J. Cuesta, Juan E. Betancort-Rijo, Stefan Gottlöber, Santiago G. Patiri, Gustavo Yepes, y Francisco Prada. Spin alignment of dark matter haloes in the shells of the largest voids. *MNRAS*, 385(2):867–874, 2008. ISSN 0035-8711. doi:10.1111/J.1365-2966.2008.12879.X. URL <https://ui.adsabs.harvard.edu/abs/2008MNRAS.385..867C/abstract>.

- L. N. da Costa y P. S. Pellegrini. The Large Scale Distribution of Galaxies in the Southern Hemisphere. En J. Audouze, M. C. Pelletan, A. Szalay, Ya. B. Zel'dovich, y P. J. E. Peebles, eds., *Large Scale Structures of the Universe*, tomo 130, pág. 121. 1988.
- Romeel Dave, Daniel Angles-Alcázar, Desika Narayanan, Qi Li, Mika H. Rafiee-rantsoa, y Sarah Appleby. SIMBA: Cosmological simulations with black hole growth and feedback. *MNRAS*, 486(2):2827–2849, 2019. doi:10.1093/mnras/stz937.
- F. Dávila-Kurbán, M. Lares, y D. Garcia Lambas. A statistical formalism for alignment analysis. *MNRAS*, en revisión, 2022. URL <http://arxiv.org/abs/2202.13244>.
- Federico Dávila-Kurbán, Ariel G. Sánchez, Marcelo Lares, y Andres N. Ruiz. Improved two-point correlation function estimates using glass-like distributions as a reference sample. *MNRAS*, 506(4):4667–4675, 2021. doi:10.1093/mnras/stab1622.
- M. Davis y P. J. E. Peebles. A survey of galaxy redshifts. V - The two-point position and velocity correlations. *ApJ*, 267:465, 1983. ISSN 0004-637X. doi:10.1086/160884.
- Marc Davis, Avery Meiksin, Michael A. Strauss, L. Nicolaci da Costa, y Amos Yahil. On the universality of the two-point galaxy correlation function. *ApJ*, 333:L9, 1988. ISSN 0004-637X. doi:10.1086/185275.
- Kyle S. Dawson, David J. Schlegel, Christopher P. Ahn, Scott F. Anderson, eric Aubourg, Stephen Bailey, Robert H. Barkhouser, et al. The Baryon Oscillation Spectroscopic Survey of SDSS-III. *AJ*, 145(1):10, 2013. doi:10.1088/0004-6256/145/1/10.
- V. de Lapparent, M. J. Geller, y J. P. Huchra. A slice of the universe. *ApJ*, 302:L1, 1986. ISSN 0004-637X. doi:10.1086/184625.
- Arnaud de Mattia y Vanina Ruhlmann-Kleider. Integral constraints in spectroscopic surveys. *JCAP*, 2019(8):036, 2019. doi:10.1088/1475-7516/2019/08/036.
- DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E. Allen, Carlos Allende Prieto, et al. The DESI Experiment Part

- I: Science, Targeting, and Survey Design. *arXiv e-prints*, arXiv:1611.00036, 2016.
- R. H. Dicke, P. J. E. Peebles, P. G. Roll, y D. T. Wilkinson. Cosmic Black-Body Radiation. *ApJ*, 142:414–419, 1965. doi:10.1086/148306.
- A. G. Doroshkevich. Spatial structure of perturbations and origin of galactic rotation in fluctuation theory. *Astrophysics*, 6(4):320–330, 1970. doi:10.1007/BF01001625.
- Y. Dubois, C. Pichon, C. Welker, D. Le Borgne, J. Devriendt, C. Laigle, S. Codis, et al. Dancing in the dark: galactic properties trace spin swings along the cosmic web. *MNRAS*, 444(2):1453–1468, 2014. doi:10.1093/mnras/stu1227.
- Frantisek Duris, Juraj Gazdarica, Iveta Gazdaricova, Lucia Strieskova, Jaroslav Budis, Jan Turna, y Tomas Szemes. Mean and variance of ratios of proportions from categories of a multinomial distribution. *Journal of Statistical Distributions and Applications*, 5, 2018. doi:10.1186/s40488-018-0083-x.
- Federico Dávila-Kurbán, Marcelo Lares, y Diego Garcia Lambas. Spin alignment around TNG300-1 voids. *MNRAS*, 2022. ISSN 0035-8711. doi:10.1093/mnras/stac3311. URL <https://doi.org/10.1093/mnras/stac3311>. Stac3311.
- eBOSS Collaboration, Shadab Alam, Marie Aubert, Santiago Avila, Christophe Balland, Julian E. Bautista, Matthew A. Bershady, et al. Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory. *PhRvD*, 103(8):083533, 2021. doi:10.1103/PhysRevD.103.083533.
- G. Efstathiou, Stephen Moody, John A. Peacock, Will J. Percival, Carlton Baugh, Joss Bland-Hawthorn, Terry Bridges, et al. Evidence for a non-zero Λ and a low matter density from a combined analysis of the 2dF Galaxy Redshift Survey and cosmic microwave background anisotropies. *MNRAS*, 330(2):L29–L35, 2002. doi:10.1046/j.1365-8711.2002.05215.x.
- Daniel J. Eisenstein, Hee-Jong Seo, Edwin Sirko, y David N. Spergel. Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak. *ApJ*, 664(2):675–679, 2007. doi:10.1086/518712.

- Daniel J Eisenstein, Idit Zehavi, David W Hogg, Roman Scoccimarro, Michael R Blanton, Robert C Nichol, Ryan Scranton, et al. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, 633(2):560–574, 2005. doi:10.1086/466512.
- Hagai El-Ad y Tsvi Piran. Voids in the Large-Scale Structure. *ApJ*, 491(2):421–435, 1997. doi:10.1086/304973.
- E. Elizalde y E. Gaztanaga. Void probability as a function of the void’s shape and scale-invariant models. *MNRAS*, 254:247–256, 1992. doi:10.1093/mnras/254.2.247.
- S. M. Fall y G. Efstathiou. Formation and rotation of disc galaxies with haloes. *MNRAS*, 193(2):189–206, 1980. ISSN 0035-8711. doi:10.1093/mnras/193.2.189.
- Hume A. Feldman, Nick Kaiser, y John A. Peacock. Power-spectrum analysis of three-dimensional redshift surveys. *ApJ*, 426:23, 1994. ISSN 0004-637X. doi:10.1086/174036.
- Karl B. Fisher, Marc Davis, Michael A. Strauss, Amos Yahil, y John P. Huchra. Clustering in the 1.2-Jy IRAS galaxy redshift survey - II. Redshift distortions and $\zeta(\text{rp}, \pi)$. *MNRAS*, 267(4):927–948, 1994. ISSN 13652966. doi:10.1093/mnras/267.4.927.
- P. Flin y W. Godlowski. The orientation of galaxies in the Local Supercluster. *MNRAS*, 222(3):525–541, 1986. ISSN 0035-8711. doi:10.1093/mnras/222.3.525.
- Jaime E. Forero-Romero, Sergio Contreras, y Nelson Padilla. Cosmic web alignments with the shape, angular momentum and peculiar velocities of dark matter haloes. *MNRAS*, 443(2):1090–1102, 2014. ISSN 0035-8711. doi:10.1093/MNRAS/STU1150. URL <https://academic.oup.com/mnras/article/443/2/1090/1056427>.
- Wendy L. Freedman y Barry F. Madore. The hubble constant. *Annual Review of Astronomy and Astrophysics*, 48(1):673–710, 2010. doi:10.1146/annurev-astro-082708-101829. URL <https://doi.org/10.1146/2Fannurev-astro-082708-101829>.
- J. N. Fry. The Galaxy correlation hierarchy in perturbation theory. *ApJ*, 279:499–510, 1984. doi:10.1086/161913.

- J. N. Fry. Cosmological density fluctuations and large-scale structure From N-point correlation functions to the probability distribution. *ApJ*, 289:10, 1985. ISSN 0004-637X. doi:10.1086/162859.
- J. N. Fry. Nonlinear statistics and biasing. *ApJ*, 308:L71, 1986. ISSN 0004-637X. doi:10.1086/184747. URL <http://adsabs.harvard.edu/doi/10.1086/184747>.
- J. N. Fry y S. Colombi. Void statistics and hierarchical scaling in the halo model. *MNRAS*, 433(1):581–590, 2013. ISSN 13652966. doi:10.1093/mnras/stt745. URL <https://ui.adsabs.harvard.edu/abs/2013MNRAS.433..581F/abstract>.
- J. N. Fry, Riccardo Giovanelli, Martha P. Haynes, Adrian L. Melott, y Robert J. Scherrer. Void Statistics, Scaling, and the Origins of Large-Scale Structure. *ApJ*, 340:11, 1989. doi:10.1086/167372.
- J. N. Fry, Riccardo Giovanelli, Martha P. Haynes, Adrian L. Melott, Robert J. Scherrer, J. N. Fry, Riccardo Giovanelli, et al. Void Statistics, Scaling, and the Origins of Large-Scale Structure. *ApJ*, 340:11, 1989. ISSN 0004-637X. doi:10.1086/167372. URL <https://ui.adsabs.harvard.edu/abs/1989ApJ...340...11F/abstract>.
- Andrea Gabrielli, Michael Joyce, y Francesco Sylos Labini. Glass-like universe: Real-space correlation properties of standard cosmological models. *PhRvD*, 65(8):083523, 2002. doi:10.1103/PhysRevD.65.083523.
- E. Gaztanaga, P. Fosalba, y E. Elizalde. Gravitational Evolution of the Large-Scale Probability Density Distribution: The Edgeworth and Gamma Expansions. *ApJ*, 539(2):522–531, 2000. doi:10.1086/309249.
- Enrique Gaztanaga, Gaztanaga, y Enrique. N-Point Correlation Functions in the CfA and SSRS Redshift Distribution of Galaxies. *ApJL*, 398:L17, 1992. ISSN 0004-637X. doi:10.1086/186566. URL <https://ui.adsabs.harvard.edu/abs/1992ApJ...398L..17G/abstract>.
- Enrique Gaztanaga y Juníchi Yokoyama. Probing the statistics of primordial fluctuations and their evolution. *ApJ*, 403:450, 1993. ISSN 0004-637X. doi:10.1086/172216.

- Margaret J. Geller y John P. Huchra. Mapping the universe. 1989. doi:10.1126/science.246.4932.897.
- Héctor Gil-Marín, Julián E Bautista, Romain Paviot, Mariana Vargas-Magaña, Sylvain de la Torre, Sebastien Fromenteau, Shadab Alam, et al. The completed SDSS-IV extended baryon oscillation spectroscopic survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0. *Monthly Notices of the Royal Astronomical Society*, 498(2):2492–2531, 2020. doi:10.1093/mnras/staa2455. URL <https://doi.org/10.1093/mnras/staa2455>.
- Daniel T Gillespie. A theorem for physicists in the theory of random variables. *American Journal of Physics*, 51:520–533, 1983. doi:10.1119/1.13221. URL <https://doi.org/10.1119/1.13221>.
- W. Godłowski. Galactic orientation within the Local Supercluster. *MNRAS*, 265(4):874–880, 1993. ISSN 0035-8711. doi:10.1093/mnras/265.4.874.
- W. Godłowski y M. Ostrowski. Investigation of galactic alignment in Local Supercluster galaxy clusters. *MNRAS*, 303(1):50–64, 1999. ISSN 0035-8711. doi:10.1046/J.1365-8711.1999.02205.X. URL <https://ui.adsabs.harvard.edu/abs/1999MNRAS.303...50G/abstract>.
- Włodzimierz Godłowski. Some aspects of the galactic orientation within the local supercluster. *MNRAS*, 271(1):19–30, 1994. ISSN 13652966. doi:10.1093/mnras/271.1.19.
- Facundo A. Gómez, Robert J. J. Grand, Antonela Monachesi, Simon D. M. White, Sebastian Bustamante, Federico Marinacci, Rüdiger Pakmor, et al. Lessons from the Auriga discs: the hunt for the Milky Way’s ex situ disc is not yet over. *MNRAS*, 472(3):3722–3733, 2017a. doi:10.1093/mnras/stx2149.
- Facundo A. Gómez, Simon D. M. White, Robert J. J. Grand, Federico Marinacci, Volker Springel, y Rüdiger Pakmor. Warps and waves in the stellar discs of the Auriga cosmological simulations. *MNRAS*, 465(3):3446–3460, 2017b. doi:10.1093/mnras/stw2957.
- III Gott, J. Richard, John Miller, Trinh X. Thuan, Stephen E. Schneider, David H. Weinberg, Charles Gammie, Kevin Polk, et al. The Topology of Large-Scale

- Structure. III. Analysis of Observations. *ApJ*, 340:625, 1989. doi:10.1086/167425.
- Jan Niklas Grieb, Ariel G. Sánchez, Salvador Salazar-Albornoz, y Claudio Dalla Vecchia. Gaussian covariance matrices for anisotropic galaxy clustering measurements. *MNRAS*, 457(2):1577–1592, 2016. ISSN 13652966. doi:10.1093/mnras/stw065.
- Oliver Hahn, C. Marcella Carollo, Cristiano Porciani, y Avishai Dekel. The evolution of dark matter halo properties in clusters, filaments, sheets and voids. *MNRAS*, 381(1):41–51, 2007. ISSN 13652966. doi:10.1111/j.1365-2966.2007.12249.x.
- A. J. S. Hamilton. Galaxy clustering and the method of voids. *ApJL*, 292:L35–L39, 1985. doi:10.1086/184468.
- A. J. S. Hamilton. Evidence for Biasing in the CfA Survey. *ApJL*, 331:L59, 1988. doi:10.1086/185235.
- A. J. S. Hamilton. Toward Better Ways to Measure the Galaxy Correlation Function. *ApJ*, 417:19, 1993a. doi:10.1086/173288.
- A. J. S. Hamilton. Linear Redshift Distortions: a Review. En Donald Hamilton, ed., *The Evolving Universe*, tomo 231 de *Astrophysics and Space Science Library*, pág. 185. 1998. doi:10.1007/978-94-011-4960-0\17.
- A. J. S. Hamilton. Formulae for growth factors in expanding universes containing matter and a cosmological constant. *MNRAS*, 322(2):419–425, 2001. doi:10.1046/j.1365-8711.2001.04137.x.
- A. J. S. Hamilton, P. Kumar, Edward Lu, y Alex Matthews. Reconstructing the Primordial Spectrum of Fluctuations of the Universe from the Observed Nonlinear Clustering of Galaxies. *ApJL*, 374:L1, 1991. doi:10.1086/186057.
- A. J. S. Hamilton. Toward Better Ways to Measure the Galaxy Correlation Function. *ApJ*, 417:19, 1993b. doi:10.1086/173288.
- Steen H. Hansen, Oscar Agertz, Michael Joyce, Joachim Stadel, Ben Moore, y Doug Potter. An Alternative to Grids and Glasses: Quaquaversal Pre-Initial Conditions for N -Body Simulations. *ApJ*, 656(2):631–635, 2007. ISSN 0004-637X. doi:10.1086/510477.

- Trevor Hastie, Robert Tibshirani, y Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.
- Alan Heavens, Alexandre Refregier, y Catherine Heymans. Intrinsic correlation of galaxy shapes: implications for weak lensing measurements. *MNRAS*, 319(2):649–656, 2000. doi:10.1046/j.1365-8711.2000.03907.x.
- Christopher M. Hirata y Uroš Seljak. Intrinsic alignment-lensing interference as a contaminant of cosmic shear. *PhRvD*, 70(6):063526, 2004. doi:10.1103/PhysRevD.70.063526.
- David W. Hogg. Distance measures in cosmology. *arXiv e-prints*, astro-ph/9905116, 1999.
- Fiona Hoyle, M. S. Vogeley, y D. Pan. Photometric properties of void galaxies in the Sloan Digital Sky Survey Data Release 7. *MNRAS*, 426(4):3041–3050, 2012. doi:10.1111/j.1365-2966.2012.21943.x.
- Fiona Hoyle y Michael S. Vogeley. Voids in the Point Source Catalogue Survey and the Updated Zwicky Catalog. *ApJ*, 566(2):641–651, 2002. doi:10.1086/338340.
- Fiona Hoyle y Michael S. Vogeley. Voids in the Two-Degree Field Galaxy Redshift Survey. *ApJ*, 607(2):751–764, 2004. doi:10.1086/386279.
- F. X. Hu, G. X. Wu, H. J. Su, y Y. Z. Liu. Morphological dependence of the orientation of the spin vectors of disk galaxies in the Virgo cluster. *A&A*, 302:45, 1995. ISSN 0004-6361. URL <https://ui.adsabs.harvard.edu/abs/1995A&A...302...45H/abstract>.
- F. X. Hu, Q. R. Yuan, H. J. Su, G. X. Wu, y Y. Z. Liu. Orientation of Spin Vectors of Bright Field Disk Galaxies in the Local Supercluster and Its Implication. *ApJ*, 495(1):179–186, 1998. ISSN 0004-637X. doi:10.1086/305267.
- E. P. Hubble. Extragalactic nebulae. *ApJ*, 64:321–369, 1926. doi:10.1086/143018.
- E. P. Hubble. *Realm of the Nebulae*. 1936.
- Edwin Hubble. A Relation between Distance and Radial Velocity among Extragalactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173, 1929. doi:10.1073/pnas.15.3.168.

- J. Huchra. Globular Clusters Associated with External Galaxies. En *Bulletin of the American Astronomical Society*, tomo 21, pág. 769. 1989.
- J. Huchra, M. Davis, D. Latham, y J. Tonry. A survey of galaxy redshifts. IV - The data. *ApJ Supplement Series*, 52:89, 1983. ISSN 0067-0049. doi:10.1086/190860.
- J. Huchra, W. Pughe, S. Kleinmann, M. Skrutski, M. Weinberg, C. Beichman, y T. Chester. The 2 Micron All Sky Survey. En Chantal Balkowski y R. C. Kraan-Korteweg, eds., *Unveiling Large-Scale Structures Behind the Milky Way*, tomo 67 de *Astronomical Society of the Pacific Conference Series*, pág. 41. 1994.
- Lluís Hurtado-Gil, Vicent J. Martínez, Pablo Arnalte-Mur, María Jesús Pons-Bordería, Cristóbal Pareja-Flores, y Silvestre Paredes. The best fit for the observed galaxy counts-in-cell distribution function. *Astronomy and Astrophysics*, 601:A40, 2017. ISSN 14320746. doi:10.1051/0004-6361/201629097.
- J. Jaaniste y E. Saar. Orientation of Spiral Galaxies as a Test of Theories of Galaxy Formation. En M. S. Longair y J. Einasto, eds., *Large Scale Structures in the Universe*, tomo 79, pág. 448. 1978.
- Y. P. Jing. Accurate Fitting Formula for the Two-Point Correlation Function of Dark Matter Halos. *ApJL*, 503(1):L9–L13, 1998. doi:10.1086/311530.
- Benjamin Joachimi, Marcello Cacciato, Thomas D. Kitching, Adrienne Leonard, Rachel Mandelbaum, Björn Malte Schäfer, Cristóbal Sifón, et al. Galaxy Alignments: An Overview. *Space Science Reviews, Volume 193, Issue 1-4, pp. 1-65*, 193(1-4):1, 2015. ISSN 0038-6308. doi:10.1007/S11214-015-0177-4. URL <https://ui.adsabs.harvard.edu/abs/2015SSRv..193....1J/abstract>.
- M. Joyce, B. Marcos, y T. Baertschiger. Towards quantitative control on discreteness error in the non-linear regime of cosmological N-body simulations. *MNRAS*, 394(2):751–773, 2009. doi:10.1111/j.1365-2966.2008.14290.x.
- F. D. Kahn y L. Woltjer. Intergalactic Matter and the Galaxy. *ApJ*, 130:705, 1959. doi:10.1086/146762.
- Nick Kaiser. Clustering in real space and in redshift space. *MNRAS*, 227:1–21, 1987. doi:10.1093/mnras/227.1.1.

- Nobunari Kashikawa y Sadanori Okamura. Spatial Orientation of Spin Vectors of Galaxies in the Local Supercluster. *PASJ*, 44:493–507, 1992. ISSN 0004-6264. URL <https://ui.adsabs.harvard.edu/abs/1992PASJ...44..493K/abstract>.
- G. Kauffmann y A. P. Fairall. Voids in the distribution of galaxies: an assessment of their significance and derivation of a void spectrum. *MNRAS*, 248:313–324, 1991. doi:10.1093/mnras/248.2.313.
- Guinevere Kauffmann, Adi Nusser, y Matthias Steinmetz. Galaxy formation and large-scale bias. *MNRAS*, 286(4):795–811, 1997. doi:10.1093/mnras/286.4.795.
- E. Keihänen, H. Kurki-Suonio, V. Lindholm, A. Viitanen, A.-S. S. Suur-Uski, V. Allevato, E. Branchini, et al. Estimating the galaxy two-point correlation function using a split random catalog. *A&A*, 631:A73, 2019. ISSN 0004-6361. doi:10.1051/0004-6361/201935828. URL <http://arxiv.org/abs/1905.01133><http://dx.doi.org/10.1051/0004-6361/201935828><https://www.aanda.org/10.1051/0004-6361/201935828><https://ui.adsabs.harvard.edu/abs/2019A%26A...631A..73K/abstract>.
- Martin Kerscher, István Szapudi, y Alexander S Szalay. A Comparison of Estimators for the Two-Point Correlation Function. *ApJL*, 535(1):L13–L16, 2000. doi:10.1086/312702.
- Alina Kiessling, Marcello Cacciato, Benjamin Joachimi, Donnacha Kirk, Thomas D Kitching, Adrienne Leonard, Rachel Mandelbaum, et al. Galaxy alignments: Theory, modelling and simulations. 2015. URL <http://www.naoj.org/Projects/HSC>;
- John R. Klauder, E. C. G. Sudarshan, y Marvin M. Miller. Highly coherent. *Physics Today*, 22(12):79, 1969. doi:10.1063/1.3035319.
- P A R Koopman. Confidence intervals for the ratio of two binomial proportions. *Biometrics*, 40:513–517, 1984. ISSN 0006341X, 15410420. URL <http://www.jstor.org/stable/2531405>.
- Katarina Kraljic, Romeel Dave, y Christophe Pichon. And yet it flips: connecting galactic spin and the cosmic web. *MNRAS*, 493(1):362–381, 2019. doi:10.1093/mnras/staa250. URL <http://arxiv.org/abs/1906.01623><http://dx.doi.org/10.1093/mnras/staa250>.

- Marc Lachieze-Rey, Luiz N. da Costa, y Sophie Maurogordato. Void Probability Function in the Southern Sky Redshift Survey. *ApJ*, 399:10, 1992. doi:10.1086/171897.
- Ofer Lahav, Makoto Itoh, Shogo Inagaki, y Yasushi Suto. Non-Gaussian Signatures from Gaussian Initial Fluctuations: Evolution of Skewness and Kurtosis from Cosmological Simulations in the Highly Nonlinear Regime. *ApJ*, 402:387, 1993. doi:10.1086/172143.
- Stephen D Landy y Alexander S Szalay. Bias and Variance of Angular Correlation Functions. *ApJ*, 412:64, 1993. doi:10.1086/172900.
- R. Laureijs, J. Amiaux, S. Arduini, J. L. Augueres, J. Brinchmann, R. Cole, M. Cropper, et al. Euclid Definition Study Report. *arXiv e-prints*, arXiv:1110.3193, 2011.
- Jounghun Lee. The Intrinsic Inclination of Galaxies Embedded in Cosmic Sheets and Its Cosmological Implications: An Analytic Calculation. *ApJL*, 614(1):L1–L4, 2004. doi:10.1086/425326.
- Jounghun Lee, Pirin Erdogdu, Jounghun Lee, y Pirin Erdogdu. The Alignments of the Galaxy Spins with the Real-Space Tidal Field Reconstructed from the 2MASS Redshift Survey. *ApJ*, 671(2):1248–1255, 2007. ISSN 0004-637X. doi:10.1086/523351. URL <https://ui.adsabs.harvard.edu/abs/2007ApJ...671.1248L/abstract>.
- Jounghun Lee y Ue-Li Pen. Cosmic Shear from Galaxy Spins. *ApJ*, Volume 532, Issue 1, pp. L5-L8., 532(1):L5, 2000. ISSN 0004-637X. doi:10.1086/312556. URL <https://ui.adsabs.harvard.edu/abs/2000ApJ...532L...5L/abstract>.
- Jounghun Lee y Ue-Li Pen. Galaxy Spin Statistics and Spin-Density Correlation. *ApJ*, Volume 555, Issue 1, pp. 106-124., 555(1):106, 2001. ISSN 0004-637X. doi:10.1086/321472. URL <https://ui.adsabs.harvard.edu/abs/2001ApJ...555...106L/abstract>.
- Jounghun Lee y Ue-Li Pen. Detection of Galaxy Spin Alignments in the Point Source Catalog Redshift Survey Shear Field. *ApJ*, Volume 567, Issue 2, pp. L111-L114., 567(2):L111, 2002. ISSN 0004-637X. doi:10.1086/340000. URL <https://ui.adsabs.harvard.edu/abs/2002ApJ...567L.111L/abstract>.

- Shihong Liao. An alternative method to generate pre-initial conditions for cosmological N-body simulations. *MNRAS*, 481(3):3750–3760, 2018. ISSN 13652966. doi:10.1093/mnras/sty2523. URL <http://arxiv.org/abs/1807.03574><http://dx.doi.org/10.1093/mnras/sty2523>.
- Noam I. Libeskind, Yehuda Hoffman, Jaime Forero-Romero, Stefan Gottlöber, Alexander Knebe, Matthias Steinmetz, y Anatoly Klypin. The velocity shear tensor: tracer of halo alignment. *MNRAS*, Volume 428, Issue 3, p.2489-2499, 428(3):2489, 2013. ISSN 0035-8711. doi:10.1093/MNRAS/STS216. URL <https://ui.adsabs.harvard.edu/abs/2013MNRAS.428.2489L/abstract>.
- Chris J. Lintott, Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick, et al. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *MNRAS*, Volume 389, Issue 3, pp. 1179-1189., 389(3):1179, 2008. ISSN 0035-8711. doi: 10.1111/J.1365-2966.2008.13689.X. URL <https://ui.adsabs.harvard.edu/abs/2008MNRAS.389.1179L/abstract>.
- Martha Lippich, Ariel G Sánchez, Manuel Colavincenzo, Emiliano Sefusatti, Pierluigi Monaco, Linda Blot, Martin Crocce, et al. Comparing approximate methods for mock catalogues and covariance matrices - I. Correlation function. *MNRAS*, 482(2):1786–1806, 2019. doi:10.1093/mnras/sty2757.
- J. Loveday, S. J. Maddox, G. Efstathiou, y B. A. Peterson. The Stromlo-APM Redshift Survey II. Variation of Galaxy Clustering with Morphology and Luminosity. *ApJ*, 442:457, 1994. doi:10.1086/175453. URL <http://arxiv.org/abs/astro-ph/9410018><http://dx.doi.org/10.1086/175453>.
- Federico Marinacci, Mark Vogelsberger, Rüdiger Pakmor, Paul Torrey, Volker Springel, Lars Hernquist, Dylan Nelson, et al. First results from the IllustrisTNG simulations: radio haloes and magnetic fields. *MNRAS*, Volume 480, Issue 4, p.5113-5139, 480(4):5113, 2018. ISSN 0035-8711. doi: 10.1093/MNRAS/STY2206. URL <https://ui.adsabs.harvard.edu/abs/2018MNRAS.480.5113M/abstract>.
- George Marsaglia, Wai Wan Tsang, y Jingbo Wang. Evaluating kolmogorov's distribution. *Journal of Statistical Software, Articles*, 8:1–4, 2003. ISSN 1548-7660. doi:10.18637/jss.v008.i18. URL <https://www.jstatsoft.org/v008/i18>.

- Takahiko Matsubara y Yasushi Suto. Scale Dependence of Three-Point Correlation Functions: Model Predictions and Redshift-Space Contamination. *ApJ*, 420:497, 1994. doi:10.1086/173580.
- Sophie Maurogordato y Marc Lachieze-Rey. Void Probabilities in the Galaxy Distribution: Scaling and Luminosity Segregation. *ApJ*, 320:13, 1987. doi:10.1086/165520.
- Sophie Maurogordato y Marc Lachieze-Rey. Void Probabilities in the Galaxy Distribution: Scaling and Luminosity Segregation. *ApJv.320, p.13*, 320:13, 1987. ISSN 0004-637X. doi:10.1086/165520. URL <https://ui.adsabs.harvard.edu/abs/1987ApJ...320...13M/abstract>.
- Aram Z. Mekjian. Generalized statistical models of voids and hierarchical structure in cosmology. *ApJ*, 655(1):1–10, 2007. doi:10.1086/508151. URL <http://arxiv.org/abs/0712.1217http://dx.doi.org/10.1086/508151>.
- H. J. Mo, Shude Mao, y Simon D. M. White. The formation of galactic discs. *MNRAS*, 295(2):319–336, 1998. doi:10.1046/j.1365-8711.1998.01227.x.
- H. J. Mo, Frank C. van den Bosch, y Simon D. M. White. Galaxy formation and evolution. *Choice Reviews Online*, 43(07):43–4006, 2006. ISSN 0009-4978. doi:10.5860/choice.43-4006. URL <https://ui.adsabs.harvard.edu/abs/2010gfe..book.....M/abstract>.
- B. Moore, T. Quinn, F. Governato, J. Stadel, y G. Lake. Cold collapse and the core catastrophe. *MNRAS*, 310(4):1147–1152, 1999. doi:10.1046/j.1365-8711.1999.03039.x.
- Jill P. Naiman, Annalisa Pillepich, Volker Springel, Enrico Ramirez-Ruiz, Paul Torrey, Mark Vogelsberger, Rüdiger Pakmor, et al. First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium. *MNRAS, Volume 477, Issue 1, p.1206-1224*, 477(1):1206, 2018. ISSN 0035-8711. doi:10.1093/MNRAS/STY618. URL <https://ui.adsabs.harvard.edu/abs/2018MNRAS.477.1206N/abstract>.
- Julio F. Navarro, Carlos S. Frenk, y Simon D. M. White. The Structure of Cold Dark Matter Halos. *ApJ*, 462:563, 1996. doi:10.1086/177173.

- Dylan Nelson, Annalisa Pillepich, Volker Springel, Rainer Weinberger, Lars Hernquist, Rüdiger Pakmor, Shy Genel, et al. First results from the IllustrisTNG simulations: the galaxy colour bimodality. *MNRAS*, Volume 475, Issue 1, p.624-647, 475(1):624, 2018. ISSN 0035-8711. doi:10.1093/MNRAS/STX3040. URL <https://ui.adsabs.harvard.edu/abs/2018MNRAS.475..624N/abstract>.
- Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, et al. The IllustrisTNG simulations: public data release. *ComAC*, 6(1):2, 2019. doi:10.1186/S40668-019-0028-X. URL <https://ui.adsabs.harvard.edu/abs/2019ComAC...6....2N/abstract>.
- Mark C. Neyrinck. ZOBOV: a parameter-free void-finding algorithm. *MNRAS*, 386(4):2101–2109, 2008. doi:10.1111/j.1365-2966.2008.13180.x.
- Peder Norberg, Carlton M. Baugh, Ed Hawkins, Steve Maddox, John A. Peacock, Shaun Cole, Carlos S. Frenk, et al. The 2dF galaxy redshift survey: Luminosity dependence of galaxy clustering. *MNRAS*, 328(1):64–70, 2001. ISSN 00358711. doi:10.1046/j.1365-8711.2001.04839.x.
- S. Otto, H. D. Politzer, J. Preskill, y M. B. Wise. The Significance of Voids. *ApJ*, 304:62, 1986. doi:10.1086/164144.
- N. D. Padilla, L. Ceccarelli, y D. G. Lambas. Spatial and dynamical properties of voids in a Λ cold dark matter universe. *MNRAS*, 363(3):977–990, 2005. doi:10.1111/J.1365-2966.2005.09500.X.
- Nikhil Padmanabhan, Xiaoying Xu, Daniel J. Eisenstein, Richard Scalzo, Antonio J. Cuesta, Kushal T. Mehta, y Eyal Kazin. A 2 per cent distance to $z = 0.35$ by reconstructing baryon acoustic oscillations - I. Methods and application to the Sloan Digital Sky Survey. *MNRAS*, 427(3):2132–2145, 2012. ISSN 00358711. doi:10.1111/j.1365-2966.2012.21888.x.
- Danny C. Pan, Michael S. Vogeley, Fiona Hoyle, Yun-Young Choi, y Changbom Park. Cosmic voids in Sloan Digital Sky Survey Data Release 7. *MNRAS*, 421(2):926–934, 2012. doi:10.1111/j.1365-2966.2011.20197.x.
- E. Panko, P. Piwowarska, J. Godłowska, W. Godłowski, y P. Flin. Orientation of galaxies in structures. *Astrophysics*, 56(3):322–331, 2013. doi:10.1007/S10511-013-9288-3.

- Aseem Paranjape, Tsz Yan Lam, y Ravi K. Sheth. A hierarchy of voids: more ado about nothing. *MNRAS*, Volume 420, Issue 2, pp. 1648-1655., 420(2):1648, 2012. ISSN 0035-8711. doi:10.1111/J.1365-2966.2011.20154.X. URL <https://ui.adsabs.harvard.edu/abs/2012MNRAS.420.1648P/abstract>.
- Changbom Park, Michael S. Vogeley, Margaret J. Geller, y John P. Huchra. Power Spectrum, Correlation Function, and Tests for Luminosity Bias in the CfA Redshift Survey. *ApJ*, 431:569, 1994. doi:10.1086/174508.
- Santiago G. Patiri, Juan E. Betancort-Rijo, Francisco Prada, Anatoly Klypin, y Stefan Gottlöber. Statistics of voids in the two-degree Field Galaxy Redshift Survey. *MNRAS*, 369(1):335–348, 2006. doi:10.1111/J.1365-2966.2006.10305.X.
- Dante Paz, Marcelo Lares, Laura Ceccarelli, Nelson Padilla, y Diego García Lambas. Clues on void evolution-II. Measuring density and velocity profiles on SDSS galaxy redshift space distortions. 436(4), 2013. doi:10.1093/MNRAS/STT1836. URL <https://ui.adsabs.harvard.edu/abs/2013MNRAS.436.3480P/abstract>.
- J. A. Peacock y R. E. Smith. Halo occupation numbers and galaxy bias. *MNRAS*, 318(4):1144–1156, 2000. doi:10.1046/j.1365-8711.2000.03779.x.
- P. J. E. Peebles. Origin of the Angular Momentum of Galaxies. *ApJ*, vol. 155, p.393, 155:393, 1969. ISSN 0004-637X. doi:10.1086/149876. URL <https://ui.adsabs.harvard.edu/abs/1969ApJ...155..393P/abstract>.
- P. J. E. Peebles y M. G. Hauser. Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs. *ApJS*, 28:19, 1974. doi:10.1086/190308.
- P. J. E. (Phillip James Edwin) Peebles. *The large-scale structure of the universe*. Princeton University Press, 1980. ISBN 9780691082400.
- A. A. Penzias y R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421, 1965. doi:10.1086/148307.
- S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, et al. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *ApJ*, 517(2):565–586, 1999. doi:10.1086/307221.

- Annalisa Pillepich, Dylan Nelson, Volker Springel, Rüdiger Pakmor, Paul Torrey, Rainer Weinberger, Mark Vogelsberger, et al. First results from the TNG50 simulation: the evolution of stellar and gaseous discs across cosmic time. *MNRAS*, Volume 490, Issue 3, p.3196-3233, 490(3):3196, 2019. ISSN 0035-8711. doi:10.1093/MNRAS/STZ2338. URL <https://ui.adsabs.harvard.edu/abs/2019MNRAS.490.3196P/abstract>.
- Annalisa Pillepich, Volker Springel, Dylan Nelson, Shy Genel, Jill Naiman, Rüdiger Pakmor, Lars Hernquist, et al. Simulating galaxy formation with the IllustrisTNG model. *MNRAS*, Volume 473, Issue 3, p.4077-4106, 473(3):4077, 2018. ISSN 0035-8711. doi:10.1093/MNRAS/STX2656. URL <https://ui.adsabs.harvard.edu/abs/2018MNRAS.473.4077P/abstract>.
- Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, et al. Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13, 2016. doi:10.1051/0004-6361/201525830.
- Manolis Plionis y Spyros Basilakos. The size and shape of local voids. *MNRAS*, 330(2):399–404, 2002. doi:10.1046/j.1365-8711.2002.05069.x.
- Beth Reid, Shirley Ho, Nikhil Padmanabhan, Will J. Percival, Jeremy Tinker, Rita Tojeiro, Martin White, et al. SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues. *MNRAS*, 455(2):1553–1573, 2016. doi:10.1093/mnras/stv2382.
- Beth A. Reid, Hee-Jong Seo, Alexie Leauthaud, Jeremy L. Tinker, y Martin White. A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies. *MNRAS*, 444(1):476–502, 2014. doi:10.1093/mnras/stu1391.
- K. Riebe, A. M. Partl, H. Enke, J. Forero-Romero, S. Gottlöber, A. Klypin, G. Lemson, et al. The multidark database: Release of the bolshoi and multidark cosmological simulations. *Astronomische Nachrichten*, 334(7):691–708, 2013. ISSN 00046337. doi:10.1002/asna.201211900.
- Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116(3):1009–1038, 1998. doi:10.1086/300499.

- Antonio Walter Riotto. Inflation and the theory of cosmological perturbations. 2002.
- Price R.M. y D G Bonett. Confidence intervals for a ratio of two independent binomial proportions. *Statistics in Medicine*, 27:5497–5508, 2008. doi:<https://doi.org/10.1002/sim.3376>.
- Andres N. Ruiz, Dante J. Paz, Marcelo Lares, Heliana E. Luparello, Laura Ceccarelli, y Diego Garcia Lambas. Clues on void evolution - III. Structure and dynamics in void shells. *MNRAS*, 448(2):1471–1482, 2015. doi:10.1093/mnras/stv019.
- Barbara S. Ryden, Adrian L. Melott, David A. Craig, III Gott, J. Richard, David H. Weinberg, Robert J. Scherrer, Suketu P. Bhavsar, et al. The Area of Isodensity Contours in Cosmological Models and Galaxy Surveys. *ApJ*, 340:647, 1989. doi:10.1086/167426.
- Ariel G. Sanchez. Arguments against using h-1 Mpc units in observational cosmology. *Physical Review D*, 102(12):123511, 2020. ISSN 24700029. doi:10.1103/PhysRevD.102.123511.
- Ariel G. Sánchez, Román Scoccimarro, Martín Crocce, Jan Niklas Grieb, Salvador Salazar-Albornoz, Claudio Dalla Vecchia, Martha Lippich, et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the configuration-space clustering wedges. *MNRAS*, 464(2):1640–1658, 2017. ISSN 13652966. doi:10.1093/mnras/stw2443.
- W. L. W. Sargent y E. L. Turner. A statistical method for determining the cosmological density parameter from the redshifts of a complete sample of galaxies. *ApJL*, 212:L3–L7, 1977. doi:10.1086/182362.
- W. C. Saslaw y A. J. S. Hamilton. Thermodynamics and galaxy clustering - Nonlinear theory of high order correlations. *ApJ*, 276:13–25, 1984. doi:10.1086/161589.
- Will Saunders, Carlos Frenk, Michael Rowan-Robinson, George Efstathiou, Andy Lawrence, Nick Kaiser, Richard Ellis, et al. The density field of the local Universe. *Nat*, 349(6304):32–38, 1991. doi:10.1038/349032a0.

- Robert J. Scherrer y Enrique Gaztanaga. The real- and redshift-space density distribution functions for large-scale structure in the spherical collapse approximation. *MNRAS*, 328(1):257–265, 2001. doi:10.1046/j.1365-8711.2001.04856.x.
- S. F. Shandarin y Ia. B. Zeldovich. Topology of the large-scale structure of the Universe. *Comments on Astrophysics*, 10:33–45, 1983.
- Sergei Shandarin, Hume A. Feldman, Katrin Heitmann, y Salman Habib. Shapes and sizes of voids in the Lambda cold dark matter universe: excursion set approach. *MNRAS*, 367(4):1629–1640, 2006. doi:10.1111/j.1365-2966.2006.10062.x.
- Ravi K. Sheth, Rien van de Weygaert, Ravi K. Sheth, y Rien van de Weygaert. A hierarchy of voids: much ado about nothing. *MNRAS*, 350(2):517–538, 2004. ISSN 0035-8711. doi:10.1111/J.1365-2966.2004.07661.X. URL <https://ui.adsabs.harvard.edu/abs/2004MNRAS.350..517S/abstract>.
- Anže Slosar y Martin White. Alignment of galaxy spins in the vicinity of voids. *Journal of Cosmology and Astroparticle Physics, Issue 06, id. 009 (2009).*, 2009(6):009, 2009. ISSN 1475-7516. doi:10.1088/1475-7516/2009/06/009. URL <https://ui.adsabs.harvard.edu/abs/2009JCAP...06..009S/abstract>.
- Thierry Sousbie. DisPerSE: robust structure identification in 2D and 3D. 2013. URL <http://arxiv.org/abs/1302.6221>.
- Volker Springel. The cosmological simulation code GADGET-2. *MNRAS*, 364(4):1105–1134, 2005. doi:10.1111/j.1365-2966.2005.09655.x.
- Volker Springel. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *MNRAS, Volume 401, Issue 2, pp. 791-851.*, 401(2):791, 2010. ISSN 0035-8711. doi:10.1111/J.1365-2966.2009.15715.X. URL <https://ui.adsabs.harvard.edu/abs/2010MNRAS.401..791S/abstract>.
- Volker Springel, Rüdiger Pakmor, Annalisa Pillepich, Rainer Weinberger, Dylan Nelson, Lars Hernquist, Mark Vogelsberger, et al. First results from the IllustrisTNG simulations: matter and galaxy clustering. *MNRAS, Volume 475, Issue 1, p.676-698*, 475(1):676, 2018. ISSN 0035-8711. doi:10.1093/MNRAS/STX3304. URL <https://ui.adsabs.harvard.edu/abs/2018MNRAS.475..676S/abstract>.

- Volker Springel, Simon D. M. White, Giuseppe Tormen, y Guinevere Kauffmann. Populating a cluster of galaxies - I. Results at $z=0$. *MNRAS*, Volume 328, Issue 3, pp. 726-750., 328(3):726, 2001. ISSN 0035-8711. doi: 10.1046/J.1365-8711.2001.04912.X. URL <https://ui.adsabs.harvard.edu/abs/2001MNRAS.328..726S/abstract>.
- Yasushi Suto y Takahiko Matsubara. Departure from Hierarchical Clustering Relations for Two-, Three-, and Four-Point Correlation Functions: Analysis of Cosmological N-Body Simulations. *ApJ*, 420:504, 1994. doi:10.1086/173581.
- István Szapudi y Alexander S. Szalay. A New Class of Estimators for the N-Point Correlations. *ApJL*, 494(1):L41–L44, 1998. doi:10.1086/311146.
- Max Tegmark, Michael R. Blanton, Michael A. Strauss, Fiona Hoyle, David Schlegel, Roman Scoccimarro, Michael S. Vogeley, et al. The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey. *ApJ*, 606(2):702–740, 2004. ISSN 0004-637X. doi:10.1086/382125. URL <http://arxiv.org/abs/astro-ph/0310725><http://dx.doi.org/10.1086/382125>.
- Elmo Tempel y Noam I. Libeskind. Galaxy Spin Alignment in Filaments and Sheets: Observational Evidence. *ApJL*, 775(2):L42, 2013. doi:10.1088/2041-8205/775/2/L42.
- Jeremy L. Tinker, Charlie Conroy, Peder Norberg, Santiago G. Patiri, David H. Weinberg, y Michael S. Warren. Void Statistics in Large Galaxy Redshift Surveys: Does Halo Occupation of Field Galaxies Depend on Environment? *ApJ*, 686(1):53–71, 2008. ISSN 0004-637X. doi:10.1086/589983.
- Jeremy L. Tinker, David H. Weinberg, y Michael S. Warren. Cosmic Voids and Galaxy Bias in the Halo Occupation Framework. *ApJ*, 647(2):737–752, 2006. ISSN 0004-637X. doi:10.1086/504795. URL <http://arxiv.org/abs/astro-ph/0603146><http://dx.doi.org/10.1086/504795>.
- Ignacio Trujillo, Conrado Carretero, y Santiago G. Patiri. Detection of the Effect of Cosmological Large-Scale Structure on the Orientation of Galaxies. *ApJ*, 640(2):L111–L114, 2006. ISSN 0004-637X. doi:10.1086/503548. URL <http://arxiv.org/abs/astro-ph/0511680><http://dx.doi.org/10.1086/503548><https://iopscience.iop.org/article/10.1086/503548>.

- Shinji Tsujikawa. Introductory review of cosmic inflation. 2003. doi:10.48550/ARXIV.HEP-PH/0304257. URL <https://arxiv.org/abs/hep-ph/0304257>.
- R. van de Weygaert y J. R. Bond. Clusters and the Theory of the Cosmic Web. En M. Plionis, O. Lopez-Cruz, y D. Hughes, eds., *A Pan-Chromatic View of Clusters of Galaxies and the Large-Scale Structure*, tomo 740, pág. 335. 2008. doi:10.1007/978-1-4020-6941-3_10.
- R. van de Weygaert y W. Schaap. The Cosmic Web: Geometric Analysis. En V.J. Martinez, E. Saar, E. Martinez-Gonzalez, y M.J. Pons-Borderia, eds., *Data Analysis in Cosmology*, tomo 665, págs. 291–413. 2009. doi:10.1007/978-3-540-44767-2_11.
- Jesús Varela, Juan Betancort-Rijo, Ignacio Trujillo, y Elena Ricciardelli. The Orientation of Disk Galaxies around Large Cosmic Voids. *ApJ*, 744(2):82, 2012. doi:10.1088/0004-637X/744/2/82.
- M. Vargas-Magaña, J. E. Bautista, J.-Ch. Hamilton, N. G. Busca, É. Aubourg, A. Labatie, J.-M. Le Goff, et al. An optimized correlation function estimator for galaxy surveys. *A&A*, 554:A131, 2013. ISSN 0004-6361. doi:10.1051/0004-6361/201220790. URL <http://www.aanda.org/10.1051/0004-6361/201220790>.
- Michael S. Vogeley, Margaret J. Geller, y John P. Huchra. Void Statistics of the CfA Redshift Survey. *ApJv.382, p.44*, 382:44, 1991. ISSN 0004-637X. doi:10.1086/170691. URL <https://ui.adsabs.harvard.edu/abs/1991ApJ...382...44V/abstract>.
- Michael S. Vogeley, Margaret J. Geller, Changbom Park, y John P. Huchra. Voids and constraints on nonlinear clustering of galaxies. *The Astronomical Journal*, 108:745, 1994. ISSN 00046256. doi:10.1086/117110.
- David H. Weinberg y Shaun Cole. Non-Gaussian fluctuations and the statistics of galaxy clustering. *MNRAS*, 259(4):652–694, 1992. doi:10.1093/mnras/259.4.652.
- Steven Weinberg. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. 1972.

- S. D. M. White. Angular momentum growth in protogalaxies. *ApJ, Part 1 (ISSN 0004-637X)*, vol. 286, Nov. 1, 1984, p. 38-41. NASA-supported research., 286:38, 1984. ISSN 0004-637X. doi:10.1086/162573. URL <https://ui.adsabs.harvard.edu/abs/1984ApJ...286...38W/abstract>.
- S. D. M. White. Large-Scale Structure. *Reviews in Modern Astronomy*, 7:255–263, 1994.
- S. D. M. White. Formation and Evolution of Galaxies. En Richard Schaeffer, Joseph Silk, Michel Spiro, y Jean Zinn-Justin, eds., *Cosmology and Large Scale Structure*, pág. 349. 1996.
- S. D. M. White y M. J. Rees. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. *MNRAS*, 183:341–358, 1978. doi:10.1093/mnras/183.3.341.
- Simon D. M. White. The hierarchy of correlation functions and its relation to other measures of galaxy clustering. *MNRAS*, 186(2):145–154, 1979. ISSN 0035-8711. doi:10.1093/mnras/186.2.145. URL <https://ui.adsabs.harvard.edu/abs/1979MNRAS.186..145W/abstract>.
- G. X. Wu. The expected isotropic distribution curves in galactic alignment. *A&A*, 452(3):807–809, 2006. ISSN 0004-6361. doi:10.1051/0004-6361:20054659. URL <https://www.aanda.org/articles/aa/abs/2006/24/aa4659-05/aa4659-05.html>.
- Renbin Yan, Martin White, y Alison L. Coil. Mock Catalogs for the DEEP2 Redshift Survey. *ApJ*, 607(2):739–750, 2004. doi:10.1086/383588.
- Abel Yang y William C. Saslaw. The Galaxy Counts-in-cells Distribution from the Sloan Digital Sky Survey. *ApJ*, 729(2):123, 2011. doi:10.1088/0004-637X/729/2/123.
- Xiaohu Yang, Frank C. van den Bosch, H. J. Mo, Shude Mao, Xi Kang, Simone M. Weinmann, Yicheng Guo, et al. The alignment between the distribution of satellites and the orientation of their central galaxy. *MNRAS*, 369(3):1293–1302, 2006. doi:10.1111/j.1365-2966.2006.10373.x.
- Donald G. York, J. Adelman, Jr. Anderson, John E., Scott F. Anderson, James Annis, Neta A. Bahcall, y J. A. etal Bakken. The Sloan Digital Sky Survey:

- Technical Summary. *AJ*, 120(3):1579–1587, 2000. ISSN 0004-6256. doi:10.1086/301513. URL <https://ui.adsabs.harvard.edu/abs/2000AJ...120.1579Y/abstract>.
- Q. R. Yuan, F. X. Hu, H. J. Su, y K. L. Huang. Orientation of the Field Galaxies in the Local Supercluster. *The Astronomical Journal*, 114:1308, 1997. ISSN 00046256. doi:10.1086/118564.
- Idit Zehavi, Michael R. Blanton, Joshua A. Frieman, David H. Weinberg, Houjun J. Mo, y Michael A. Strauss. Galaxy Clustering in Early SDSS Redshift Data. *ApJ*, 571(1):172–190, 2001. doi:10.1086/339893. URL <http://arxiv.org/abs/astro-ph/0106476><http://dx.doi.org/10.1086/339893>.
- Idit Zehavi, Zheng Zheng, David H. Weinberg, Michael R. Blanton, Neta A. Bahcall, Andreas A. Berlind, Jon Brinkmann, et al. Galaxy clustering in the completed sdss redshift survey: The dependence on color and luminosity. *ApJ*, 736(1):59, 2011. ISSN 15384357. doi:10.1088/0004-637X/736/1/59.
- Ya. B. Zel’dovich. Gravitational instability: An approximate theory for large density perturbations. *A&A*, 5:84–89, 1970.
- Youcai Zhang, Xiaohu Yang, Huiyuan Wang, Lei Wang, Wentao Luo, H. J. Mo, y Frank C. van den Bosch. Spin Alignments of Spiral Galaxies within the Large-scale Structure from SDSS DR7. *ApJ*, 798(1):17, 2015. doi:10.1088/0004-637X/798/1/17.
- Qirong Zhu, Federico Marinacci, Moupiya Maji, Yuexing Li, Volker Springel, y Lars Hernquist. Baryonic impact on the dark matter distribution in Milky Way-sized galaxies and their satellites. *MNRAS*, 458(2):1559–1580, 2016. doi:10.1093/mnras/stw374.
- F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, 1933.