Università degli Studi di Napoli "Federico II"

Scuola Politecnica e delle Scienze di Base
Area Didattica di Scienze Matematiche Fisiche e Naturali

Dipartimento di Fisica "Ettore Pancini"

Laurea triennale in Fisica

Stime della funzione di correlazione a due punti per un cluster di galassie

Relatori:
Prof. Salvatore Capozziello
Dr.ssa Micol Benetti

Candidato:
Alberto Tortora
Matricola N85000597

A.A. 2017/2018

Indice

Capitolo 1

Introduzione

1.1 Il Principio Cosmologico

La nostra principale fonte di informazioni riguardo la struttura dell'Universo è la radiazione elettromagnetica proveniente dai corpi luminosi che lo popolano; l'uomo, quindi, ha sempre cercato di analizzare lo spettro elettromagnetico che giungeva fino a lui. Per prima cosa gli studiosi hanno preso in considerazione la porzione di spettro che comprendeva il visibile, sia ad occhio nudo che con l'aiuto di telescopi. Galileo Galilei, astronomo, fisico, matematico e filosofo italiano, nonché padre fondatore del metodo scientifico, fu il primo studioso ad utilizzare il telescopio rifrattore per uso astronomico a Venezia nel 1609. Si deve a lui, infatti, il perfezionamento di questo strumento. Con lo sviluppo della tecnologia, in seguito, si sono prese in considerazione anche altre bande dello spettro elettromagnetico. In primis le *onde radio*, raccolte per la prima volta nel 1933 dall'antenna di Karl Jansky, fondatore della **radioastronomia**, che studiò per primo la radiazione emessa dalla nostra galassia (la Via Lattea). In seguito i *raggi-X*, scoperti nel 1962 da un'equipe guidata da Riccardo Giacconi, astrofisico italiano, co-vincitore del Premio Nobel per la fisica nel 2002. Poco dopo vennero osservati *raggi* γ, provenienti dallo spazio ed assorbiti dall'atmosfera terrestre e, di conseguenza, non rilevabili senza l'ausilio di veicoli spaziali. Il primo telescopio per l'osservazione dei raggi gamma fu mandato in orbita a bordo del satellite *Explorer 11* nel 1961. Abbiamo anche rilevato, infine, *onde gravitazionali*, previste nel 1916 nell'ambito della teoria della relatività generale, ma confermate empiricamente soltanto 100 anni dopo, l'11 febbraio 2016, grazie allo studio sulla fusione di due buchi neri distanti circa 1 miliardo e 300 milioni di anni luce.

Le stelle che osserviamo si organizzano in strutture più ampie chiamate *galassie* che a loro volta si raggruppano in *ammassi* più grandi e questi ultimi si organizzano, ancora una volta, in *superammassi*, separati da giganteschi vuoti. Questa struttura dell'Universo è quella che chiamiamo **cosmo a grande scala**. Quando rivolgiamo la nostra attenzione su di esso, appare subito evidente che una descrizione della geometria che lo caratterizza non può essere ricavata basandosi soltanto sulle osservazioni. Questo è dovuto anche alla velocità finita della luce nel vuoto che, dovendo giungere da remote regioni dell'Universo, impiegherà un tempo finito permettendoci di aver accesso soltanto ad informazioni riguardanti il passato, sempre più remoto man mano che spingiamo la nostra attenzione più lontano. Per comprendere quanto in là stiamo "guardando nel passato" si necessita, quindi, di strumenti per conoscere la distanza da cui ha avuto inizio la radiazione. Uno degli strumenti di cui si è in possesso è proprio l'analisi dello spostamento verso il rosso delle linee spettrali, il cosiddetto *redshift*. Esso è il fenomeno che si osserva quando una radiazione elettromagnetica emessa da un oggetto ha una lunghezza d'onda maggiore rispetto a quella che aveva all'emissione. Nel caso della luce, il colore si sposta nella direzione del rosso, l'estremo inferiore dello spettro del visibile, da qui il nome. In generale, che la radiazione elettromagnetica sia visibile o meno, un redshift significa un aumento della lunghezza d'onda, equivalente ad una diminuzione della frequenza o ad una minore energia dei fotoni. Dato che lo spazio percorso dalla radiazione è legato sia all'ipotizzata espansione dell'Universo, sia al movimento mutuale dei corpi nel cosmo, l'analisi delle righe spettrali ci fornisce quantitativamente la distanza che ha percorso quella determinata radiazione elettromagnetica. Il nostro *cono di luce* è, dunque, limitato ad una piccola regione dell'Universo e per questo motivo è inevitabile che un modello cosmologico riguardante il cosmo a grande scala sia basato su ipotesi teoriche accettabili, oltre che sulle osservazioni.

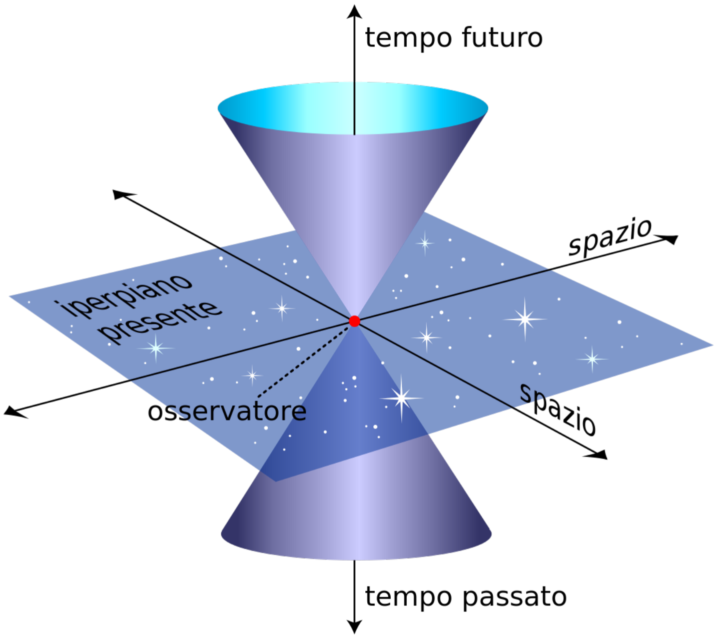

Figura 1.1: Il cono di luce è una porzione nei diagrammi dello spazio-tempo di Minkowski dove sono confinati tutti gli eventuali sviluppi futuri e le possibili cause nel passato di un evento attuale, situato nell'origine dello spazio-tempo, al vertice del cono. Il *principio di causalità* stabilisce che ciò che accade al centro del cono di luce (il qui ed ora) può influenzare solo ciò che accadrà nel cono di luce futuro e può essere influenzato solo da ciò che è avvenuto nel cono di luce passato.

La cosmologia moderna è basata sul cosiddetto Principio Cosmologico: l'ipotesi in cui si afferma che l'Universo a grandi scale è spazialmente omogeneo ed isotropo. Esso si basa su una generalizzazione del Principio Cosmologico Copernicano che rovesciò la convinzione che la Terra si trovasse al centro dell'Universo, affermando che fosse il Sole ad essere il centro. Il nostro pianeta, dunque, non si trova in una posizione privilegiata nell'Universo, ma in una posizione generica, perfettamente equivalente ad un'altra. L'estensione progressiva del principio Copernicano ci porta quindi ad affermare che un osservatore, su scala cosmica, osserverebbe ovunque la stessa distribuzione di materia ed energia. Se ci limitassimo ad osservare la regione più vicina a noi, sulla scala delle decine di migliaia di anni luce, ciò che vedremmo sarebbe unicamente la nostra galassia, la quale ci appare decisamente non omogenea con le stelle che si concentrano principalmente sul piano galattico. Allargando la nostra visuale sulla scala di milioni di anni luce, scopriamo che le galassie possono essere considerate come corpi isolati, una dall'altra, da milioni di anni luce. Anche su questa scala, quindi, possiamo notare che siamo ben lontani da considerare il cosmo isotropo ed omogeneo. In questo contesto è d'uopo ricordare il nome di Edwin Powell Hubble, astronomo che classificò per primo alcune strutture del cosmo come oggetti esterni alla nostra galassia (1925). A lui si deve, soltanto 4 anni dopo, anche lo studio della dinamica dell'Universo e la scoperta che quest'ultimo si stia espandendo secondo la legge :

$$v = H_0\, d$$

dove v è la velocità di allontanamento della galassia nella direzione della nostra linea di vista, d è la distanza della galassia dalla Terra e H_0 è una costante di proporzionalità detta "costante di Hubble". Ma in che modo, quindi, possiamo affermare che l'Universo è isotropo ed omogeneo? Con le moderne tecnologie siamo in grado di allargare ancora maggiormente questa scala, scoprendo che le galassie si aggregano in ammassi che si estendono per centinaia di milioni di anni luce intervallati da regioni di vuoto del medesimo ordine di grandezza. È sulla scala dei miliardi di anni luce, dunque, che l'Universo ci appare isotropo in tutte le direzioni ed è in questa scala che possiamo affermare che l'omogeneità comincia ad essere visibile. Un'altra conferma dell'isotropia e dell'omogeneità dell'Universo arrivò con la scoperta della radiazione cosmica di fondo (CMB), nel 1964, grazie allo straordinario studio degli astronomi statunitensi **Arno Penzias** e **Robert Woodrow Wilson** che li portò a conseguire il Premio Nobel per la fisica nel 1978. La CMB è stata poi misurata da svariati satelliti[1] in ogni direzione, pressoché isotropa se non per fluttuazioni di ordini di grandezza di $10^{-5}\mu K$. Lo spettro della radiazione di fondo corrisponde allo spettro di radiazione di un corpo nero con una temperatura $T_0 = (2,7280 \pm 0,004)K$[2]. Rappresentando il più perfetto modello di corpo nero che si sia mai ritrovato in natura, la CMB è considerata una prova consistente dell'ipotesi cosmogonica del Big Bang[3] e fornisce rilevanti informazioni sulla struttura dell'Universo primordiale, essendo una radiazione emessa in tempi antichissimi, quando il nostro Universo era molto più piccolo di adesso. Esaminando questa radiazione si è constatato che l'Universo è composto, ad una prima osservazione, da materia visibile (detta materia *barionica*) e da radiazione. Uno dei dilemmi che ha attanagliato gli scienziati del secolo scorso fu proprio questo: un Universo così composto

[1]Ricordiamo tra tutti il *Cosmic Background Explorer* (COBE - lanciato dalla NASA nel 1989) ed i più moderni *WMAP* (2001) e *Planck Surveyor* (2009).

[2]Temperatura attuale della CMB, scesa di più di 1000 volte rispetto al momento della creazione dei primi atomi, per tale motivo si trova il pedice *0* in T_0.

[3]Il Big Bang è un modello cosmologico basato sull'idea che l'universo iniziò a espandersi a velocità elevatissima a partire da una condizione iniziale di temperatura e densità estreme (*singolarità*) e che questo processo continui tuttora. È il modello predominante nella comunità scientifica e ha avuto conferme basate su diverse osservazioni astronomiche.

dovrebbe contrarsi e collassare per auto-gravitazione e non espandersi. Si è rilevato, inoltre, che le galassie presentano una velocità di rotazione molto maggiore rispetto alla quantità di materia osservata e dunque, per descrivere gli oggetti compatti che osserviamo, le galassie non possono essere formate solo da materia visibile, ma devono essere più massive. Per far quadrare i conti, quindi, gli astronomi ipotizzano la presenza di una materia *oscura* (che fornisce una massa maggiore alle galassie) ed energia *oscura* (che bilancia la forza gravitazionale ed è responsabile dell'espansione osservata dell'Universo). Nonostante la loro natura sia ancora sconosciuta ed ampiamente discussa in svariati modelli cosmologici, si pensa che ad esse sia legato circa il 95 % di densità di energia dell'Universo. Al momento, il modello cosmologico più popolare è il Λ-CDM model, un unverso piatto in cui circa il 75 % della densità di energia è dovuta ad una costante cosmologica (Λ), il 21 % alla materia oscura "fredda" (CDM) ed il restante 4 % alla materia barionica di cui stelle e galassie sono costituite.

1.2 La formazione di strutture a larga scala

Per spiegare la presenza di strutture, in particolare di galassie, abbiamo chiaramente bisogno di alcune deviazioni dalla perfetta uniformità ed isotropia. La cosmologia standard, però, non fornisce di per sé una spiegazione per l'origine di queste perturbazioni. Si pensa che questo particolare stato di isotropia ed omogeneità del cosmo sia mutato molto presto, quando l'Universo era così denso che gli effetti quantistici avevano una certa valenza. Una conseguenza possibile di tale ipotesi è la generazione di perturbazioni di densità dovute a fluttuazioni quantiche durante il periodo dell'*inflazione*[4]. Queste fluttuazioni, seppur presenti in scala microscopica, hanno avuto ripercussioni gigantesche a causa del processo di inflazione che ingrandendo lo spazio microscopico a dimensioni cosmiche avrebbe dato origine a piccole disomogeneità gravitazionalmente instabili. Tale ipotesi viene anche confermata da osservazioni di disomogeneità in temperatura sulla mappa del CMB. Queste fluttuazioni di temperatura, infatti, indicano regioni sovradense e regioni sottodense, rispettivamente evolute poi (ed oggi osservabili) in regioni dove la materia si è aggregata ed in regioni di vuoto. L'idea che le strutture possano essersi

[4] l'inflazione è una teoria che ipotizza che l'Universo, poco dopo il Big Bang, abbia attraversato una fase di espansione estremamente rapida, dovuta a una grande pressione negativa. Proposta da Alan Guth nel 1981 e revisionato nel'82 da Linde, Albrecht e Steinhardt, si stima che sia avvenuta intorno a 10^{-35} secondi dal Big Bang, sia durata per $\sim 10^{-30}$ secondi ed abbia aumentato il raggio dell'universo di un fattore enorme, tra 10^{25} e 10^{30}

formate attraverso un'instabilità gravitazionale in questo modo ha origine da Jeans (1902), che ha dimostrato che la stabilità di una perturbazione dipende dalla competizione tra gravità e pressione. Esattamente, le instabilità crescono se la forza di auto-gravitazione per unità di massa, F_g, diventa maggiore della forza per unità di massa data dalla pressione, F_p:

$$F_g \simeq \frac{GM}{\lambda^2} \simeq \frac{G\rho\lambda^3}{\lambda^2} > F_p \simeq \frac{p\lambda^2}{\rho\lambda^3} \simeq \frac{v_s^2}{\lambda} \tag{1.1}$$

con v_s velocità del suono. Dalla relazione 1.1 segue che la crescita della perturbazione avviene per $\lambda_j > v_s(G\rho)^{-1/2}$. Le perturbazioni di densità, dunque, crescono solo se sono più grandi (più pesanti) di una lunghezza (massa) caratteristica ora indicata come *Lunghezza di Jeans* λ_j (massa di Jeans), oltre la quale la gravità è in grado di superare i gradienti di pressione. L'applicazione di questo criterio ad un background in espansione è stata elaborata, tra gli altri, da Gamow e Teller (1939) e in seguito da Lifshitz (1946), con il risultato che la crescita delle perturbazioni avvenisse in potenza nel tempo, piuttosto che esponenziale come avveniva in un contesto statico. Durante gli anni '60 e '70 si è compreso che la cosmologia standard presentava diverse gravi carenze, una fra tutte era che le diverse parti dell'Universo che vediamo oggi non siano mai state in contatto causale nell'Universo primordiale. Come possono allora queste regioni essere così simili, come richiesto dall'isotropia della CMB? La risposta è da ricercare ancora una volta nel periodo dell'inflazione. In questo scenario diverse parti dell'Universo odierno erano effettivamente in contatto causale prima che avvenisse l'inflazione, consentendo in tal modo ai processi fisici di stabilire l'omogeneità e l'isotropia. In molti modelli inflazionari, le fluttuazioni quantistiche dell'energia del vuoto possono produrre perturbazioni di densità con proprietà coerenti con la struttura su larga scala osservata oggi. L'inflazione offre quindi una spiegazione promettente per l'origine fisica delle perturbazioni iniziali. Sfortunatamente, la nostra comprensione dell'universo primordiale è ancora lontana dall'essere completa, di conseguenza, anche questa parte della teoria della formazione delle galassie è ancora parzialmente fenomenologica: le condizioni iniziali tipicamente sono specificate da un insieme di parametri che sono vincolati da dati osservativi, come lo schema delle fluttuazioni nella radiazione cosmica di fondo o l'abbondanza odierna di ammassi di galassie.

1.3 Ammassi di Galassie

Le galassie si raggruppano in strutture più o meno dense. Le strutture con popolazione meno densa di galassie sono chiamate comunemente "gruppi", mentre gli aggregati con una densità più elevata, circa 50 membri in un volume di $\sim$ 50 Megaparsec di estensione, sono chiamati "ammassi di galassie" (Galaxy clusters). Gruppi ed Ammassi sono le strutture più massive dell'Universo (da 10^{14} $M_{\odot}$ a 10^{15} $M_{\odot}$) e sono importanti per lo studio dell'evoluzione della popolazione delle galassie. Grazie alla loro grande densità ed alla loro luminosità, queste strutture sono visibili da grandi distanze diventando delle vere e proprie sonde cosmologiche. Per distinguere un semplice raggruppamento di galassie da un cluster o da un gruppo, ai fini statistici, si è soliti utilizzare due criteri di selezione, attraverso i quali Abell nel 1958 catalogò 1682 ammassi:

- *Richness criterion*: ogni ammasso deve avere almeno 50 galassie con magnitudine apparente $m < m_3+2$, dove m_3 è la magnitudine apparente del terzo membro più luminoso. La "ricchezza" di un cluster, quindi, è definita come il numero di galassie membri con magnitudine apparente tra m_3 e m_3 $+2$. I cluster di Abell sono quelli con una ricchezza maggiore di 50, nonostante abbia elencato cluster con ricchezza nell'intervallo da 30 a 50.

- *Compactness criterion*: solo i membri con distanze dal centro del cluster inferiori a 1,5 h^{-1}Mpc (raggio di Abell) sono selezionate come membri del cluster. Dato il criterio di ricchezza, il criterio di compattezza è equivalente ad un criterio di densità.

Abell, inoltre, classificò i cluster come *regolari* ed *irregolari*. I primi sono chiamati regolari perché caratterizzati da una distribuzione di galassie concentrata e da una simmetria circolare. I due cluster più studiati, a causa della loro vicinanza, sono l'Ammasso della Vergine e l'Ammasso dellla Chioma. Il primo si presenta irregolare e ricco di sottostrutture; il secondo, al contrario, è regolare, molto massivo e con una grande simmetria. I cluster sono in generale ricchi di galassie ellittiche (E) o lenticolari (S0). La presenza di strutture di tipo E + S0 è $\sim$80% nei cluster regolari e $\sim$50% nei cluster irregolari. Probabilmente, quindi, le galassie subiscono trasformazioni morfologiche in ambienti densi.
Nonostante la differenza tra un gruppo ed un cluster non sia proprio netta, possiamo affermare che un gruppo, tipicamente, è un sistema formato da 3 a 30 galassie, ha un'estensione che può variare da 0.1 a 1 h^{-1}Mpc ed una massa di $10^{12.5}$ - 10^{14} $M_{\odot}$. Il gruppo più studiato è il gruppo di cui facciamo parte, *Il Gruppo Locale*,con un'estensione di circa 1 Mpc, formato principalmente da due galassie: la *Nebulosa di Andromeda* e la Via Lattea con le sue *piccole* galassie satelliti, le nubi di Magellano.

1.4 La distribuzione spaziale di strutture a larga scala

Una delle proprietà più importanti della popolazione di una galassia, o di un cluster, è la distribuzione spaziale. Essendo formate da una gran quantità di massa, ci aspettiamo che la loro distribuzione rifletta in qualche modo la distribuzione di massa della struttura a larga scala dell'Universo. Per questo motivo, uno studio dettagliato della distribuzione spaziale può contenere informazioni non soltanto riguardanti la distribuzione della materia, connessa fortemente alla cosmologia, ma anche importanti indizi sulla loro formazione. Come accennato, le galassie non si distribuiscono in modo casuale nello spazio, ma mostrano una grande varietà di strutture, più o meno popolate e dense. La maggior parte delle galassie si riunisce in enormi addensamenti piatti, come dei giganteschi "fogli", ed altri allungati, detti "filamenti", separati tra loro da immense regioni vuote (*voids*), di circa 100 Mpc di diametro, contenenti poche o nessuna galassia.
Uno dei problemi della cosmologia è che le strutture che sottoponiamo alla nostra attenzione sono oggetti tridimensionali, mentre le immagini che raccogliamo sono necessariamente basate sulla proiezione delle posizioni della galassia sulla nostra sfera celeste, pertanto, gli effetti della proiezione sono inevitabili. Sovrastime casuali della densità di popolazione sulla sfera celeste, a causa della proiezione delle strutture osservate a più profondità, possono facilmente essere classificate come cluster. Allo stesso modo è possibile l'effetto inverso: a causa delle fluttuazioni della densità numerica delle galassie in primo piano, un cluster ad alto redshift può essere classificato come una fluttuazione insignificante e, quindi, non essere rivelato. Non a caso, non tutti i membri di un cluster, classificati come tali, si sono rivelati essere galassie. Gli effetti della proiezione giocano un ruolo importante, soprattutto se la stima del redshift è alquanto approssimata. Analisi spettroscopiche eseguite per molti cluster di Abell, infatti, hanno presentato un errore sulla sua stima dei redshift di $\sim 30\%$, stima sorprendentemente accurata se si tiene conto che il suo catalogo si basava su ispezioni visive di lastre fotografiche.
La maggior parte dei cluster regolari presenta una densità di galassie che aumenta progressivamente verso il centro. Se il cluster non è molto ellittico, questa distribuzione di densità può essere assunta, in prima approssimazione, come sfericamente simmetrica. Solo la distribuzione di densità proiettata N(R) è osservabile. Questo valore è correlato alla densità numerica tridimensionale n(r) attraverso la formula:

$$N(R) = \int_{-\infty}^{\infty} dz\, n\,(\sqrt{R^2 + z^2}) = 2\int_{R}^{\infty} \frac{dr\, r\, n(r)}{\sqrt{r^2 - R^2}} \tag{1.2}$$

dove nel secondo integrale si è fatta una trasformazione della variabile di integrazione z (linea di vista) al raggio in tre dimensioni $r = \sqrt{R^2 + z^2}$.

Di solito si considerano forme parametrizzate di N(R) e si adattano i parametri alle posizioni delle galassie osservate. Una distribuzione parametrizzata ha bisogno di almeno 5 parametri per descrivere le caratteristiche basiliari di un cluster:

- Due parametri descrivono il centro del cluster sulla sfera celeste;
- un parametro per descrivere l'ampiezza della densità, ad esempio la densità centrale $N_0 = N(0)$;
- un parametro descrive la scala caratteristica di un cluster. Spesso è utilizzato il *core radius* r_c, definito come il raggio a cui la densità proiettata equivale alla metà della densità centrale, ovvero $N(r_c) = \frac{N_0}{2}$;
- un ultimo parametro indica la fine del cluster, in prima approssimazione si sceglie il raggio di Abell.

1.5 Proprietà statistiche come strumento per studiare il cosmo

Al momento nessun modello dell'Universo è in grado di descrivere la distribuzione di materia prossima alla Via Lattea nel dettaglio. Non ci sono modelli fisici in grado di predire che ad una certa distanza dalla nostra galassia ci sia una seconda galassia massiva con una certa forma, poiché queste caratteristiche dipendono dalle specifiche condizioni iniziali della distribuzione della materia nell'Universo primordiale. Si possono, però, provare a studiare le proprietà statistiche di una determinata distribuzione di massa, come per esempio, la densità numerica media di ammassi di galassie al di sopra di una data massa o la probabilità di trovare una galassia massiva ad una certa distanza da un'altra. Con le simulazioni, dunque, è possibile generare modelli cosmologici che abbiano le stesse proprietà statistiche del nostro Universo. Essendo stato osservato che le galassie si raggruppano in grandi strutture, sembra ragionevole affermare che la probabilità di trovare una galassia in corrispondenza di una posizione x non è indipendente dal fatto che ci sia un'altra galassia in prossimità di x. In altre parole, è più probabile trovare una galassia in prossimità di un altra che in una posizione arbitraria casuale. Questo fenomeno è descritto dalle cosiddette *funzioni di correlazione*, di cui si discuterà nel prossimo capitolo.

Capitolo 2

La funzione di correlazione

2.1 La statistica delle galassie

Qui si vuole studiare in che modo si può esaminare la formazione di strutture osservando aggregamenti di galassie. Prima di procedere nell'uso di una particolare statistica, bisogna essere sicuri che il campione scelto per misurare quantitativamente il raggruppamento di galassie (*clustering*) nel nostro Universo sia grande abbastanza da essere, in un certo senso, rappresentativo dell'Universo stesso, ovvero con un *piano di campionamento* sulla scala dei cluster di galassie. Su scale cosmiche maggiori, invece, la statistica è molto meno efficace e lo studio si basa su dei principi primi considerato che le osservazioni permettono di misurare con accettabile precisione solo la porzione più prossima dell'Universo. Lo studio di regioni più lontane infatti è soggetto ad errori sistematici (*bias*) giacché vengono rilevati solo gli oggetti più luminosi. Alcuni studiosi quindi suggerirono vari modelli cosmologici tra cui alcuni in cui la distribuzione della materia nell'universo non fosse omogenea, ma seguisse una *legge frattale*[1]. In ogni modo, un effetto da considerare su queste scale è il seguente: una volta effettuate misure statistiche su un campione finito, il risultato così ottenuto sarebbe differente nel caso si prendesse lo stesso campione delle stesse dimensioni, ma situato in un altro luogo del nostro Universo. Questo è l'effetto generalmente conosciuto come "*varianza cosmica*" e definisce l'incertezza statistica inerente alle osservazioni del cosmo a larga scala, poiché si ipotizza che il piano di campionamento sia troppo piccolo rispetto alla scala dell'intero Universo.
Uno strumento statistico molto utilizzato in cosmologia è quello delle *funzio-*

[1]Un Universo frattale è un modello cosmologico ipotizzante un Universo che si ripete nella sua forma ed allo stesso modo su scale diverse; ingrandendo una qualunque sua parte si ottiene dunque una figura simile all'originale. Uno di questi modelli fu proposto dall'astronomo franco-statunitense **Gérard de Vaucouleurs** negli anni '70.

ni di correlazione che sono state per molti anni il metodo più comune per descrivere il raggruppamento di galassie o di ammassi. Proposte già negli anni '60 da **Totsuji** e **Kihara**, esse hanno trovato in **Phillip James Edwin "Jim" Peebles** il loro più grande sostenitore. Quest'ultimo infatti, insieme a diversi colleghi, realizzò un programma durante gli anni '70 in grado di stimare queste funzioni e tanti altri set di dati dal *Lick Galaxy Catalogue*[2]. Questa strumento può essere inoltre utilizzato non solo per le varie tipologie di galassie (nella scala del visibile, dell' infrarosso, radio etc.), ma anche per strutture più complesse come quasar o cluster.

2.2 Funzione di correlazione spaziale ed angolare

La funzione di correlazione rappresenta *l'eccesso di probabilità*, rispetto a quella descritta da una statistica di Poisson, di rivelare una coppia di galassie con una distanza mutuale $\boldsymbol{r}$. Tramite questa funzione, in altre parole, possiamo stimare quanto è probabile che le galassie siano più o meno aggregate rispetto ad un insieme di galassie casuale, distribuite in modo omogeneo. Questo insieme può essere tridimensionale (utilizzato nello studio della *fuzione di correlazione spaziale*), ma si possono ottenere risultati utili anche da una distribuzione di oggetti a due dimensioni, ovvero, da una distribuzione di punti proiettati sulla sfera celeste (*funzione di correlazione angolare*). Si considerino due punti e due elementi di volume dV intorno questi punti. Se chiamiamo n_v la densità numerica media delle galassie nello spazio, allora la probabilità di trovare una galassia nell'elemento volume dV attorno alla posizione x sarà data da:

$$dP_1 = n_v dV$$

tale probabilità risulta indipendente da x se assumiamo che l'Universo sia statisticamente omogeneo. Per rendere trascurabile la probabilità di trovare più galassie nello stesso elemento di volume, si usa assumere $\frac{dV}{V} \ll 1$. La probabilità congiunta di trovare una galassia nell'elemento di volume dV_1 e nello stesso tempo trovare un'altra galassia nell'elemento di volume dV_2, separate da un vettore $\boldsymbol{r_{12}}$ è quindi:

$$d^2P_2 = n_v^2[1 + \xi(r_{12})]dV_1 dV_2 \tag{2.1}$$

[2]Il Lick Galaxy Catalogue, anche conosciuto come "Il catalogo di Shane e Wirtanen", è rinomato per avere al suo interno all'incirca un milione di galassie compilate interamente a occhio, senza l'aiuto della digitalizzazione da computer.

dove la funzione $\xi(r)$ è chiamata *funzione di correlazione a due punti spaziale*. Da notare che la funzione ξ dipende soltanto dal modulo del vettore $\boldsymbol{r_{12}}$ e non dalla sua direzione, per la presenza delle ipotesi di isotropia ed omogeneità. Essa è definita in termini della probabilità congiunta di trovare due galassie separate da una distanza una distanza definita rispetto alla probabilità che ci si aspetterebbe per una distribuzione casuale. Se le galassie fossero sparse in modo completamente casuale nello spazio, allora si avrebbe chiaramente $\xi(r_{12}) \equiv 0$. Con una $\xi(r_{12}) > 0$, le galassie risultano raggruppate, in caso contrario ($\xi(r_{12}) < 0$) esse non presentano alcuna correlazione e quindi tenderebbero ad evitarsi. Nel caso in cui la funzione di correlazione risultasse positiva, essa ad ogni modo dovrà diventare negativa su larga scala (con $r_{12} \to \infty$), sia perché sappiamo che le grandi strutture a larga scala si allontanano tra loro ($\xi < 0$) come visto nel paragrafo 1.1, sia per far sì che converga l'integrale sul volume della 2.1. Bisogna, però, sottolineare che la funzione di correlazione a due punti evolve molto rapidamente nel regime non lineare.

L'equazione 2.1 implica che il numero medio di galassie ad una certa distanza r da una galassia data sia:

$$\langle N \rangle_r = \frac{4}{3}\pi n_v r^3 + 4\pi n_v + \int_0^r \xi(r'_{12})\, r'^{\,2}_{12}\, dr'_{12}$$

dove il secondo termine rappresenta il numero in eccesso paragonato ad una distribuzione uniforme.

Solitamente poiché le osservazioni sono proiettate sulla sfera celeste, la funzione di correlazione viene calcolata dalle posizioni angolari degli oggetti nei cataloghi, ma naturalmente se si hanno a disposizione i redshift degli oggetti si può calcolare direttamente la funzione di correlazione di tipo spaziale. Con ipotesi e assunzioni aggiuntive si può stimare poi la funzione spaziale da quella angolare attraverso l'*equazione di Limber* che verrà in seguito discussa. Da un catalogo bidimensionale è possibile definire la *funzione di correlazione a due punti angolare,* $w(\theta)$ attraverso la formula:

$$d^2P_2 = n_\Omega^2[1 + w(\theta_{12})]d\Omega_1 d\Omega_2 \tag{2.2}$$

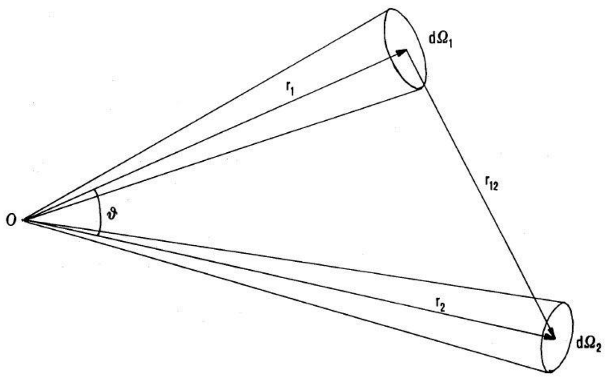

Figura 2.1: In analogia con la 2.1, l'equazione 2.2 rappresenta la probabilità di trovare due galassie in due piccoli elementi di angolo solido $d\Omega_1$ e $d\Omega_2$, separati da un angolo θ_{12} sulla sfera celeste, mentre n_Ω rappresenta il numero medio di galassie calcolato per unità di angolo solido.

2.3 L'equazione di Limber

Uno degli aspetti più importanti della funzione di correlazione a due punti è la possibilità di scrivere una relazione tra la versione angolare e quella spaziale. Fu **Limber** nel 1954 ad introdurre questa relazione e da lui prende nome la sua equazione. Egli parte dalla *funzione di luminosità*, $\phi(L)$, per definire il numero di galassie per unità di volume con una luminosità compresa nel range $L + dL$:

$$dN = \phi(L)dL$$

Questa funzione viene poi convertita in una funzione della *magnitudine*[3] $\psi(M)$ che permette di scrivere:

$$d^2P = \psi(M)dMdV \tag{2.3}$$

[3]La magnitudine può essere apparente o assoluta. La magnitudine apparente (m) di un corpo celeste è una misura della sua luminosità rilevabile dalla Terra. Il valore della magnitudine è corretto in modo da ottenere la luminosità che l'oggetto avrebbe se la Terra fosse priva di atmosfera. La magnitudine assoluta (M) è la magnitudine apparente che un oggetto avrebbe se si trovasse ad una distanza dall'osservatore di 10 parsec (o 1 unità astronomica) a seconda del tipo di oggetto (oggetto stellare o del Sistema solare). É definita in modo tale che più un oggetto è intrinsecamente luminoso, più la sua magnitudine assoluta è numericamente bassa, anche negativa. Ogni grado della scala corrisponde ad un incremento (o decremento) di circa $2,512$ volte. In altre parole, una stella che presenta magnitudine $+1$ è circa $2,512$ volte più luminosa di una che presenti $+2$ come magnitudine.

che, analogamente alla funzione di luminosità, definisce la probabilità di trovare una galassia con una magnitudine apparente compresa tra M e $M + dM$ in un volume dV. Analogamente all'equazione 2.1 possiamo scrivere la probabilità congiunta di trovare due galassie, una in un elemento di volume dV_1 con una magnitudine compresa tra M_1 e $M_1 + dM_1$ ed un'altra in dV_2 con magnitudine compresa tra M_2 e $M_2 + dM_2$, separate da una distanza $\boldsymbol{r_{12}}$:

$$d^4P = [\psi(M_1)\psi(M_2) + G(M_1, M_2, r_{12})]dM_1 dM_2 dV_1 dV_2 \tag{2.4}$$

dove G tiene conto della correlazione tra le due galassie. Limber suppone (*Ipotesi di Limber*) che la magnitudine assoluta di una galassia sia statisticamente indipendente dalla posizione che essa assume rispetto alle altre, in altre parole, che la luminosità non varia in funzione della densità dell'ambiente circostante. Questo ci permette di esplicitare la G in questo modo:

$$G(M_1, M_2, r_{12}) = \psi(M_1)\psi(M_2)\xi(r_{12}) \tag{2.5}$$

Nella compilazione di un catalogo, sono raccolte tutte le galassie che sono più luminose rispetto ad una magnitudine definita m_0 in una regione della sfera celeste circoscritta. Si effettueranno, naturalmente, degli errori sistematici per gli oggetti con magnitudine $m \simeq m_0$ e per tale motivo diventa necessario introdurre una funzione di selezione $f(m - m_0)$ che rappresenta la probabilità di includere una galassia con una magnitudine apparente m. Teoricamente un catalogo dovrebbe presentare un taglio netto al valore $m = m_0$, ma è ovviamente molto difficile da realizzare nella pratica. La funzione di luminosità delle galassie ha una magnitudine caratteristica dopo la quale, per stelle più luminose e quindi per magnitudine minore, tende molto rapidamente a zero. Il valore è $M^* \simeq -19,5 + 5\log h$. Data la definizione di magnitudine assoluta:

$$M = m + 5 - 5\log_{10} d$$

si può calcolare la distanza tipica a cui si deve trovare un osservatore per notare una magnitudine apparente $m = M^*$:

$$D^* = 10^{0.2(m_o - M^*) - 5}\text{Mpc}. \tag{2.6}$$

Il numero di galassie per unità di angolo solido all'interno di un catalogo, per la 2.3 e la 2.6, equivale a:

$$n_\Omega = D^{*3} \int_0^\infty x^2 dx \int_{-\infty}^\infty \psi(M) f(M - M^* + 5\log x) dM = D^{*3} \int_0^\infty \Psi(x)\, x^2 dx \tag{2.7}$$

dove $x = \frac{r}{D^*}$ e

$$\Psi(x) = \int_{-\infty}^{\infty} \psi(M) f(M - M^* + 5\log x)\, dM$$

La funzione $\Psi(x)$ rappresenta il numero di galassie per unità di volume ad una distanza data $r = xD^*$, appartenente al catalogo. Essa è espressa con buona approssimazione con una funzione potenza di x.
Confrontando la 2.2 con la 2.4 e tenendo conto della 2.5, è possibile scrivere:

$$\begin{aligned} d^2P_2 &= n_\Omega^2[1 + w(\theta_{12})]d\Omega_1 d\Omega_2 \\ &= D^{*6}\int_0^{\infty} \Psi(x_1)x_1^2 dx_1 \int_0^{\infty} \Psi(x_2)x_2^2[1 + \xi(r_{12})]dx_2 d\Omega_1 d\Omega_2 \end{aligned} \tag{2.8}$$

dove

$$r_{12}^2 = D^{*2}(x_1^2 + x_2^2 - 2x_1x_2\cos\theta_{12}) \tag{2.9}$$

É conveniente ora introdurre due nuove variabili:

$$x = \frac{1}{2}(x_1 + x_2), \qquad y = \frac{x_1 - x_2}{x^2\theta_{12}}\,.$$

Sottolineando che la funzione di luminosità tende subito a zero per $M < M^*$, si può affermare che la scala della lunghezza delle correlazioni debba essere molto minore di D^* o al più uguale. Ricordando che $x_i = r_i/D^*$, i maggior contributi di $\xi(r_{12})$ nella 2.8 verranno dai termini con $x_1 \simeq x_2 \simeq 1$, separati da un piccolo angolo θ_{12}. Dunque, con le variabili introdotte si ha:

$$y^2x^2\theta_{12}^2 = (x_1 - x_2)^2$$

$$x^2 = \frac{x_1^2 + x_2^2 - 2x_1x_2}{4} + x_1x_2 \quad\Rightarrow\quad x^2\left(1 - \frac{y^2\theta_{12}^2}{4}\right) = x_1x_2$$

Per piccoli angoli si può scrivere $\cos\theta_{12} \approx 1 - \frac{\theta_{12}^2}{2}$, l'equazione 2.9 diventa:

$$\begin{aligned} r_{12}^2 &= D^{*2}(x_1^2 + x_2^2 - 2x_1x_2 + x_1x_2\theta_{12}^2) \\ r_{12}^2 &= D^{*2}\left(y^2x^2\theta_{12}^2 + x^2\theta_{12}^2 - \frac{x^2y^2\theta_{12}^4}{4}\right) \\ r_{12}^2 &\simeq D^{*2}x^2\theta_{12}^2(1 + y^2) \end{aligned} \tag{2.10}$$

avendo considerato θ^4 trascurabile rispetto a θ^2. Quest'ultima equazione, insieme alla 2.8, trascurando termini in θ di grado superiore al primo, permette di scrivere **l'equazione di Limber**:

$$w(\theta_{12}) \simeq \frac{\theta_{12} \int_0^\infty \Psi^2(x) x^5 dx \int_{-\infty}^\infty \xi[D^* x \theta_{12}(1+y^2)^{1/2}] dy}{[\int_0^\infty \Psi(x) x^2 dx]^2} \tag{2.11}$$

che fornisce un'importante relazione tra la funzione di correlazione di tipo spaziale (ξ) ed angolare (w). Proprietà interessante di questa equazione è la possibilità di scegliere un angolo $\theta'_{12} = \frac{D^*}{D'^*}\theta_{12}$ tale che si abbia:

$$w'(\theta'_{12}) = \frac{D^*}{D'^*} w(\theta_{12})$$

ottenendo così una corrispondenza tra w e w', due funzioni di correlazione di due cataloghi con distanza caratteristica D^* e D'^*.

2.4 Stimare la funzione di correlazione

Da un qualunque survey, non è possibile ricavare in maniera immediata la funzione di correlazione nello spazio tridimensionale. Non è possibile, infatti, ricavare direttamente la distanza mutuale delle galassie mediante il loro redshift, ma è possibile determinare la loro posizione in termini di coordinate angolari sulla volta celeste e di distanza in redshift lungo la linea di vista. Bisogna considerare che nella valutazione del redshift si presentano delle distorsioni dovute agli effetti dei vari moti. La velocità apparente, infatti, risulta essere:

$$v_a = v_H + v_p$$

dove v_a è la velocità apparente fornita dal redshift, v_H la velocità dovuta al moto complessivo di espansione dell'Universo (velocità di Hubble) e v_p la proiezione lungo la linea di vista del moto peculiare della galassia (moto rispetto al sistema di riferimento a cui appartiene). Per questi motivi si potrebbe stimare in primis la funzione di correlazione a due punti di tipo angolare ed in seguito ricavare quella spaziale tramite l'equazione di Limber (2.11). Nel corso degli anni, ad ogni modo, diversi stimatori sono stati proposti per ovviare a questi problemi e misurare la funzione di correlazione con la minima distorsione e varianza possibile (**Peebles** nel 1980 e più accuratamente **Landy e Szalay** nel 1993).

Dall'analisi dei cataloghi bidimensionali proiettati sulla sfera celeste (Lick Galaxy Catalogue, ma anche i survey più recenti, *APM* (1990) e *COSMOS*

(2006)), si evince che, in un appropriato intervallo dell'angolo θ, la funzione di correlazione angolare $w(\theta)$ può essere ben approssimata dalla legge di potenza:

$$w(\theta) \simeq A\theta^{-\delta} \qquad (\theta_{min} \leq \theta \leq \theta_{max};\ \delta \simeq 0.8) \tag{2.12}$$

dove l'ampiezza A dipende dalla distanza caratteristica D^* delle galassie nel catalogo. L'intervallo angolare a cui si riferisce la 2.12 corrisponde ad una separazione spaziale $0.1\,h^{-1}\mathrm{Mpc} \leq r \leq 10\,h^{-1}\mathrm{Mpc}$ a questa distanza. Il regime lineare in cui vale la 2.12 è quell'intervallo angolare in cui, graficando la funzione di correlazione (w) in funzione dell'angolo θ, entrambe in scala logaritmica, ed effettuando una regressione lineare si otterrebbe una pendenza numericamente uguale a $-\delta$; oltre questo regime la funzione di correlazione angolare tende ad annullarsi molto velocemente. Assumiamo, dunque, che anche la correlazione a due punti spaziale, in una scala appropriata, sia data da:

$$\xi(r) = Br^{-\gamma} \tag{2.13}$$

Allora dalla 2.11 si può immediatamente ricavare:

$$w(\theta) = C\theta^{1-\gamma} = C\theta^{-\delta}$$

dove C è una costante legata alla B tramite la stessa 2.11. É ragionevole scrivere, continuando dalla 2.13:

$$\xi(r) = (\frac{r}{r_{0g}})^{-\gamma} \tag{2.14}$$

con $r_{0g} \simeq 5\,h^{-1}\mathrm{Mpc}$ e $\gamma \simeq 1.8$ nel range $0.1\,h^{-1}\mathrm{Mpc} \leq r \leq 10\,h^{-1}\mathrm{Mpc}$. Su scale maggiori, anche la funzione di correlazione spaziale tende rapidamente a zero. La quantità r_{0g}, laddove $\xi = 1$, è chiamata *lunghezza di correlazione*. Per ricavare questi dati empiricamente è necessario analizzare un generico survey di galassie ed eseguire un fit, per farlo abbiamo bisogno di stimatori[4]. Partendo dalla definizione data dalla 2.1, è possibile valutare la funzione di correlazione confrontando il numero di coppie di un catalogo di oggetti reali dati con il numero di coppie contenute in un catalogo random completamente omogeneo, generato con una geometria identica al primo da una distribuzione *Poissoniana*. Siano n ed N_R il numero di galassie presenti rispettivamente in un catalogo dato ed in un catalogo omogeneo casuale e siano $DD(r)$, $RR(r)$ e $DR(r)$ il numero di coppie formate accoppiando oggetti (a distanza mutuale r)

[4]Uno stimatore è una funzione che associa ad ogni elemento di un campione di dati un valore del parametro da stimare. In altre parole, è una funzione di un campione di dati estratti casualmente da una popolazione.

appartenenti rispettivamente entrambi al catalogo dato, entrambi al catalogo casuale o uno al catalogo dato ed un altro al catalogo casuale. Siano, infine, N_{DD}, N_{RR} e N_{DR} il numero totale di coppie nel catalogo (reali, casuale o ibride). Con la convenzione di contare le coppie soltanto una volta, si ottiene:

$$\begin{aligned} N_{DD} &= \frac{n(n-1)}{2} \\ N_{RR} &= \frac{n(N_R-1)}{2} \\ N_{DR} &= nN_R \end{aligned}$$

Si possono ottenere così degli stimatori della funzione di correlazione, di cui i più comuni sono stati introdotti da **Peebles & Hauser** (1974), **Davis & Peebles** (1983), **Hamilton** (1993), **Landy & Szalay** (1993):

$$\xi_{PH} = \frac{N_{RR}}{N_{DD}}\frac{DD(r)}{RR(r)} - 1 \tag{2.15}$$

$$\xi_{DP} = \frac{N_{DR}}{N_{DD}}\frac{DD(r)}{DR(r)} - 1 \tag{2.16}$$

$$\xi_{H} = \frac{N_{DR}^2}{N_{DD}N_{RR}}\frac{DD(r)RR(r)}{[DR(r)]^2} - 1 \tag{2.17}$$

$$\xi_{LS} = 1 + \frac{N_{RR}}{N_{DD}}\frac{DD(r)}{RR(r)} - 2\frac{N_{RR}}{N_{DR}}\frac{DR(r)}{RR(r)} \tag{2.18}$$

Come possiamo notare, il più accurato è quello di Landy-Szalay che considera anche le coppie ad incrocio (*cross-pair*) tra il catalogo reale osservato ed il catalogo omogeneo generato in modo random. Nel caso in cui il catalogo osservato e quello casuale abbiano lo stesso numero di oggetti ($N_{DD} = N_{RR} = N_{DR}$) lo stimatore diventa:

$$\xi_{LS} = 1 + \frac{DD(r)}{RR(r)} - 2\frac{DR(r)}{RR(r)} = \frac{RR(r) + DD(r) - 2DR(r)}{RR(r)}$$

Questo stimatore, rispetto a quello utilizzato da Peebles, risente di meno delle fluttuazioni e delle incertezze dovute al numero finito di campioni presi in esame e la sua varianza è quasi del tutto *Poissoniana*, ovvero, dipende solo dalla radice del numero di coppie del catalogo reale dato $\sqrt{N_{DD}}$. Proprio per la sua accuratezza, lo stimatore più utilizzato per il calcolo della funzione di correlazione a due punti è quello di Landy-Szalay anche se ciò comporta inevitabilmente un incremento del tempo di calcolo dovuto alla necessità di valutare anche le coppie miste tra catalogo e distribuzione casuale.

2.5 Funzioni di correlazione di ordine superiore

Fin qui si è descritto come costruire una statistica per due galassie separate da una distanza mutuale $\boldsymbol{r}$ all'interno di una distribuzione di galassie. Per far ciò è stata considerata solamente la funzione di correlazione a due punti che descrive, come è stato illustrato, l'eccesso o il difetto di probabilità di trovare due galassie alla distanza indicata rispetto ad una distribuzione omogena. Tuttavia, per descrivere completamente le caratteristiche di una distribuzione di N oggetti, come un cluster di galassie, è necessario conoscere tutte le funzioni di correlazione fino all'ordine $N-1$. Potremmo definire l'eccesso di probabilità di trovare una struttura di N punti separati da un vettore di distanze $\boldsymbol{r} = \boldsymbol{r_{ij}}$ con $(i,j) \in [1,n] \times [1,n]$ in maniera analoga all'equazione 2.1 da:

$$dP_m = n_v^m[1 + \xi_m(\boldsymbol{r})]dV_1 dV_2 ... dV_m \tag{2.19}$$

La funzione ξ_m è detta *funzione di correlazione di ordine* m e contiene tutti i contributi di ordine $n < m$. Valendo la 2.14 entro certe scale, si ipotizza che per ordini superiori si possa applicare un *modello gerarchico* in cui la funzione all'N-esimo ordine sia collegata alla funzione di correlazione di ordine $N-1$ ed in questo modo, iterando il processo, ricondursi sempre al caso di ordine 2, generalizzando quanto esposto finora. Tuttavia, le funzioni di correlazione di ordine superiore al secondo, essendo ricavate su tutte le distanze che separano gli N punti considerati, sono molto più complicate da interpretare fisicamente. Supponendo che la scala caratteristica non vari, possiamo considerare la correlazione a tre punti:

$$dP_3 = n_v^3[1 + \xi_{1,2} + \xi_{2,3} + \xi_{1,3} + \zeta_{1,2,3}]dV_1 dV_2 dV_3 \tag{2.20}$$

dove $\xi_{i,j} = \xi(r_{ij})$ e $\zeta_{i,j,k} = \zeta(r_{i,j}, r_{j,k}, r_{i,k})$ che, in base alle assunzioni fatte, può essere riscritta in questo modo:

$$\zeta_{1,2,3} = Q(\xi_{1,2}\xi_{2,3} + \xi_{2,3}\xi_{3,1} + \xi_{3,1}\xi_{1,2})$$

dove Q è una costante opportuna. Questa forma approssima i dati osservati in modo appropriato con un valore di $Q \simeq 1$ e nel range $50\,h^{-1}\text{kpc} < r < 5\,h^{-1}\text{Mpc}$. La generalizzazione dell'equazione precedente per le funzioni di correlazione di ogni ordine risulta essere:

$$\xi_{(N)} = \sum_{Topologie} Q_{N,t} \sum_{Permutazioni} \prod_{Edges} \xi_{i,j} \tag{2.21}$$

La formula 2.21 mostra che per avere una funzione di correlazione di ordine superiore al secondo, bisogna effettuare una produttoria tra le funzioni di correlazione degli $N-1$ punti considerati, sommare su tutte le possibili permutazioni di indici con i quali denotiamo gli oggetti e sommare nuovamente il tutto su tutte le possibili topologie (disposizioni) che si formano connettendo gli N oggetti considerati. Nonostante questo modello abbia riscosso un notevole successo, bisogna sottolineare che le informazioni statistiche contenute in queste funzioni sono limitate. Bisognerebbe conoscere le funzioni di correlazione di tutti gli ordini per avere una descrizione statistica completa delle proprietà di una distribuzione di punti, ma questo nella pratica è complicato da realizzare. Non bisogna dimenticare, inoltre, che le funzioni di correlazione su larga scala, anche quelle ad ordini più basse, sono in generale molto difficili da determinare e mancano di una trattazione intuitiva. A scale ancor maggiori,infatti, oltre i 10 Mpc h^{-1}, la funzione di correlazione, seppur rispettando la 2.14, tende ad assumere valori negativi suggerendo un'analisi matematica basata più su principi primi e modelli cosmologici a larga scala che sul calcolo delle funzioni di correlazione, in questo caso molto approssimato.

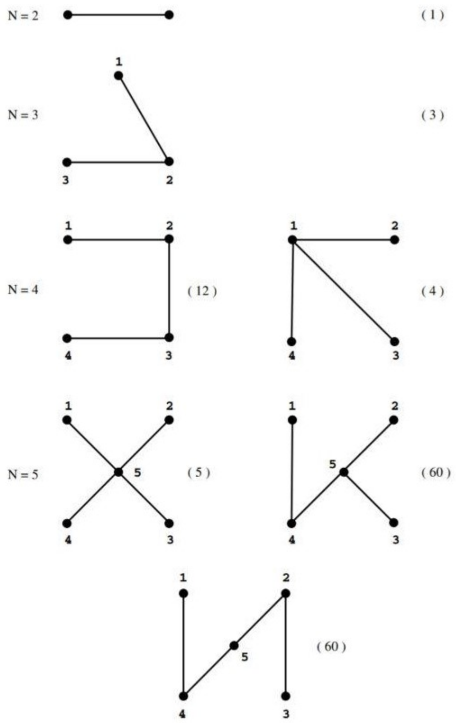

Figura 2.2: Rappresentazioni di differenti topologie esistenti nel modello gerarchico connettendo N punti. In particolare sono mostrate le topologie per $N=1, N=2$ e $N=3$. Il valore mostrato in parentesi sulla destra indica il numero di permutazioni possibili.

Capitolo 3

Confronto delle varianze degli stimatori

Fin qui si è presentata una statistica per caratterizzare il clustering delle galassie e sono stati proposti diversi stimatori per il calcolo della funzione di correlazione a due punti. Generalmente viene ipotizzato che la varianza negli stimatori sia quella di una statistica Poissoniana dei conteggi in ogni intervallo (Peebles, 1980), ma questa è solo una previsione ottimistica che ad un'analisi approfondita, proposta nel seguito, si rivela essere vera solo per due degli stimatori proposti nel capitolo precedente (lo stimatore di Hamilton e di Landy&Szalay). In seguito si propone una discussione sulla varianza della funzione di correlazione a due punti di tipo angolare ma, data l'equazione 2.11, la trattazione è equivalente nel caso di quella spaziale. I problemi maggiori che si riscontrano nella stima della funzione di correlazione sono dovuti alle dimensioni finite dello spazio campionato ed ai vincoli tra i campioni e i confini dello spazio di campionamento. Si è visto che gli stimatori nello specifico si basano su tre quantità (DD, DR e RR) le quali sono variabili casuali caratterizzate da due aspetti principali: dipendono dal numero di galassie esaminate nel campione (n) e da come esse sono distribuite nello spazio preso in considerazione. Nel seguito scriveremo $\langle A \rangle$ per indicare la media di una generica quantità A su tutto lo spazio.
Il valore atteso della quantità RR è proporzionale alla frazione delle coppie totali presente negli intervalli a distanza $\theta \pm d\frac{\theta}{2}$ in una distribuzione uniforme e può essere calcolata, con metodi computazionali, con un grado di precisione arbitraria [Landy&Szalay, 1993].

3.1 Le fluttuazioni statistiche sul valore atteso

Nel calcolo del bias[1] e della varianza è utile esprimere le tre quantità citate in funzione del loro valore atteso e delle sue fluttuazioni:

$$\begin{aligned} DD &= \langle DD \rangle (1+\alpha) \\ DR &= \langle DR \rangle (1+\beta) \\ RR &= \langle RR \rangle (1+\gamma), \end{aligned}$$

dove α, β e γ sono le fluttuazioni sul valore atteso rispettivamente delle quantità DD, DR e RR. Per un gran numero di dati random (N_R) e per il motivo discusso nel paragrafo precedente, si considera la varianza di γ arbitrariamente piccola e che le fluttuazioni in DR siano dovute solo alla variazione dei punti del catalogo di dati reali. Con questo formalismo, fermandosi al secondo ordine per le fluttuazioni, la media e la varianza dello stimatore 2.16 di tipo angolare, sarà:

$$\begin{aligned} 1+w_2(\theta) &= \left\langle \frac{DD}{DR} \right\rangle = \frac{2N_R}{n-1} \frac{\langle DD \rangle}{\langle DR \rangle} \left\langle \frac{1+\alpha}{1+\beta} \right\rangle \\ \text{con } \langle \alpha \rangle = \langle \beta \rangle &= 0 \text{ per definizione e } \beta \text{ molto piccolo} \Rightarrow \frac{1}{1+\beta} \simeq 1-\beta+\beta^2 \\ &\simeq \frac{2N_R}{n-1} \left\langle \frac{DD}{DR} \right\rangle (1 - \langle \alpha\beta \rangle + \langle \beta^2 \rangle), \\ var[w_2(\theta)] &\simeq \left(\frac{2N_R}{n-1} \frac{\langle DD \rangle}{\langle DR \rangle} \right)^2 (\langle \alpha^2 \rangle + \langle \beta^2 \rangle - 2\langle \alpha\beta \rangle). \end{aligned} \tag{3.1}$$

Il calcolo del bias e della varianza della fuzione di correlazione per ogni stimatore si lega al calcolo dei coefficienti sopra indicati. Normalizzando le fluttuazioni si ha:

$$\begin{aligned} \langle \alpha^2 \rangle &= \frac{\langle DD \cdot DD \rangle - \langle DD \rangle^2}{\langle DD \rangle^2}, \\ \langle \beta^2 \rangle &= \frac{\langle DR \cdot DR \rangle - \langle DR \rangle^2}{\langle DR \rangle^2}, \\ \langle \alpha\beta \rangle &= \frac{\langle DD \cdot DR \rangle - \langle DD \rangle \langle DR \rangle}{\langle DD \rangle \langle DR \rangle}. \end{aligned} \tag{3.2}$$

[1] Il bias è uno scostamento o una distorsione dal valore atteso stimato. Un campione distorto, in generale, fornisce una stima falsata delle caratteristiche della popolazione studiata.

3.1.1 Conteggio di coppie $\langle DD \rangle$ non correlate

Si supponga di avere n punti distribuiti omogeneamente in un area Ω e si divida tale area in K celle. Si ipotizzi, inoltre, di aumentare il numero di celle affinché in ogni cella ci sia un numero di galassie pari a 0 o 1. La probabilità di trovare un oggetto in ogni cella sarà quindi data da $\langle v \rangle = \frac{n}{K}$ e ci si aspetta che per una seconda cella essa sia $\langle v' \rangle = \frac{(n-1)}{K-1}$. Nel caso in cui i punti non siano correlati, la probabilità congiunta $\langle v_i v_j \rangle$ che entrambe le celle contengano un punto è semplicemente il prodotto tra le due probabilità.

Nel caso non correlato, dunque, il valore atteso di coppie con una separazione angolare di θ sarà:

$$\langle DD \rangle = \left\langle \sum_{i<j}^{K} v_i v_j \Theta_{ij}^{\theta} \right\rangle = \sum_{i<j}^{K} \langle v_i v_j \rangle \Theta_{ij}^{\theta} \tag{3.3}$$

dove $\Theta_{ij}^{\theta} = 1$ se le celle i, j sono separate da una distanza $\theta \pm d\frac{\theta}{2}$ o zero altrimenti. Definito $G_p(\theta)$, un fattore di forma pari alla frazione di coppie di celle separate da una distanza $\theta \pm d\frac{\theta}{2}$, si può scrivere :

$$\sum_{i<j}^{K} \Theta_{ij}^{\theta} = \frac{K(K-1)}{2} G_p(\theta) \tag{3.4}$$

Il valore atteso delle coppie di celle in un campione di n galassie diventa:

$$\langle DD \rangle = \frac{n(n-1)}{K(k-1)} G_p(\theta) \frac{K(K-1)}{2} = \frac{n(n-1)}{2} G_p(\theta) \tag{3.5}$$

dove, in particolare, è utile notare come il valore atteso sia indipendete da K. Sia $P(n)$ la probabilità di ottenere n galassie da una distribuzione Poissoniana con valore atteso N, allora si potrebbe stimare la media di insieme del numero di coppie:

$$E_N(DD) = \sum_{n=2}^{\infty} \langle DD \rangle P(n) = G_p(\theta) \sum_{n=2}^{\infty} \frac{n(n-1)}{2} P(n) = \frac{N^2}{2} G_p(\theta).$$

Il termine $\langle DD \cdot DD \rangle$ sarà dato da:

$$\langle DD \cdot DD \rangle = \left\langle \sum_{i<j}^{K} v_i v_j \Theta_{ij}^{\theta} \sum_{k<l}^{K} v_k v_l \Theta_{kl}^{\theta} \right\rangle = \sum_{i<j}^{K} \sum_{k<l}^{K} \langle v_i v_j v_k v_l \rangle \Theta_{ij}^{\theta} \Theta_{kl}^{\theta} \tag{3.6}$$

Ove, essendo le v_i uguali a zero o uno in ogni cella, anche il loro valore atteso sarà $v = \langle v_i \rangle = \langle v_i^2 \rangle$. Possiamo concentrarci allora sulla sommatoria delle Θ nella 3.6 dividendola in una sommatoria di 3 termini relativi alle degenerazioni esistenti. Nel caso in cui tutti gli indici siano diversi tra loro, le combinazioni possibili sono:

$$\binom{K}{2}\binom{K-2}{2} = \frac{K(K-1)(K-2)(K-3)}{4}.$$

Indicando con $G_q(\theta)$ la frazione di celle contenenti coppie di punti i, j e l, k separate mutualmente da una distanza $\theta \pm d\frac{\theta}{2}$, la sommatoria relativa a questo termine diventa:

$$\sum_{i<j}^{K}\sum_{k<l}^{K}\Theta_{ij}^{\theta}\Theta_{kl}^{\theta} = \frac{K(K-1)(K-2)(K-3)}{4}G_q(\theta)$$

Nel caso in cui uno dei 4 indici si sovrapponga, si ha una degenerazione tra due indici, ovvero un punto posizionato in una cella centrale e due punti a distanza $\theta \pm d\frac{\theta}{2}$ dal primo. Sia in questo caso $G_t(\theta)$ la frazione di tripletti di celle, la sommatoria relativa a questo secondo termine diventa:

$$\sum_{i<j}^{K}\sum_{k<l}^{K}\Theta_{ij}^{\theta}\Theta_{kl}^{\theta} = K(K-1)(K-2)G_t(\theta).$$

L'ultima sommatoria relativo al fattore di forma $G_p(\theta)$ racchiude la frazione di conteggi relativi ai casi di doppia degenerazione, in cui i, j e k, l sono sovrapposti. Considerando che il numero totale dei termini non nulli nella somma deve essere uguale al quadrato del numero di coppie, si può scrivere:

$$\begin{aligned}\left[G_p(\theta)\frac{K(K-1)}{2}\right]^2 =& G_q(\theta)\frac{K(K-1)(K-2)(K-3)}{4}+ \\ &+ G_t(\theta)K(K-1)(K-2) + G_p(\theta)\frac{K(K-1)}{2}.\end{aligned} \tag{3.7}$$

Da quest'ultima è possibile ricavare $G_q(\theta) = G_p(\theta)^2$, dividendola per K^4 e considerando un numero di celle molto ampio. Moltiplicando ogni termine per la propria probabilità, si eliminano ancora una volta le dipendenze da K, così come verificato nell'equazione 3.5, sicché il valore atteso dal quadrato del numero DD equivale a:

$$\begin{aligned}\langle DD \cdot DD\rangle =& \frac{n(n-1)(n-2)(n-3)}{4}G_p^2(\theta)+ \\ &+ n(n-1)(n-2)G_t(\theta) + \frac{n(n-1)}{2}G_p(\theta).\end{aligned} \tag{3.8}$$

e la sua varianza sarà dunque:

$$\begin{aligned} Var(DD) =& \frac{n(n-1)(n-2)(n-3)}{4} G_p^2(\theta) - \left[G_p(\theta) \frac{n(n-1)}{2} \right]^2 + \\ &+ n(n-1)(n-2)G_t(\theta). \end{aligned} \tag{3.9}$$

Se si andasse a considerare la varianza quando n è distribuita con andamento Poissoniano di valore atteso N, la varianza sulla popolazione diventa:

$$var_N(DD) = N^3 G_t(\theta) + \frac{N^2}{2} G_p(\theta),$$

dove il termine $N^3 G_t(\theta)$, dipende dal quadrato della larghezza dell'intervallo e dunque solo per un gran numero di celle con una larghezza infinitesima la varianza della popolazione tende alla varianza Poissoniana. Si può notare, infatti, che l'equazione 3.9 sovrastima di molto la varianza nel caso in cui n sia relativamente piccolo. Con l'equazione 3.9 e la 3.5 si può inoltre ricavare il valore medio del quadrato della fluttuazione α:

$$\langle \alpha^2 \rangle = \frac{2}{n(n-1)} \left\{ 2(n-2) \left[\frac{G_t(\theta)}{G_p^2(\theta)} - 1 \right] + \frac{1}{G_p(\theta)} - 1 \right\}. \tag{3.10}$$

3.1.2 Conteggio di coppie $\langle DD \rangle$ correlate

Nel caso in cui ci fosse correlazione tra i punti presi in considerazione, il calcolo della varianza deve tener conto della funzione di correlazione esistente. Definiamo la media della funzione di correlazione a due punti sull'intero spazio di campionamento:

$$w_\Omega = \int_\Omega G_p(\theta) w(\theta) d\Omega. \tag{3.11}$$

Tenendo conto della correlazione, il valore atteso del prodotto $\langle v_i v_j \rangle$ sarà:

$$\langle v_i v_j \rangle = \frac{n(n-1)}{K(K-1)} [1 + w(\theta)].$$

L'inclusione della funzione di correlazione rende il prodotto $\langle DD \rangle$ non normalizzato, per questo motivo si aggiunge un fattore moltiplicativo che si dimostra essere:

$$\frac{1}{1 + w_\Omega}.$$

Il valore atteso del numero di coppie di galassie risulta:

$$\langle DD \rangle = \frac{n(n-1)}{2} G_p(\theta) \frac{1+w(\theta)}{1+w_\Omega}. \tag{3.12}$$

Prendendo in considerazione l'equazione 3.8, si modificano tutti i termini secondo la funzione di correlazione tenendo conto dell'eccesso di probabilità di trovare quadrupletti, tripletti e coppie (Landy&Szalay, 1993). Il primo termine indica la probabilità di trovare due coppie i cui membri sono a distanza $\theta \pm d\frac{\theta}{2}$ tra loro. Nell'approssimazione di deboli correlazioni, il primo termine diventa:

$$\frac{n(n-1)(n-2)(n-3)}{4} G_p^2(\theta) \frac{1+2w(\theta)}{1+2w_\Omega}.$$

Allo stesso modo, il secondo termine che indica la probabilità di trovare tripletti, ovvero tre punti di cui due ad una distanza $\theta \pm d\frac{\theta}{2}$ da un terzo, diventa:

$$n(n-1)(n-2) G_t(\theta) \frac{1+3w(\theta)}{1+3w_\Omega}.$$

L'ultimo termine è semplicemente il valore atteso del numero di coppie a distanza $\theta \pm d\frac{\theta}{2}$ ricavato dalla funzione di correlazione (equazione 3.12). Sommando questi tre termini si ricava il valore atteso del quadrato del numero di coppie $\langle DD \cdot DD \rangle$ e con l'equazione 3.12 il valore di $\langle \alpha^2 \rangle$:

$$\langle \alpha^2 \rangle = \frac{2}{n(n-1)} \left\{ 2(n-2) \left[\frac{G_t(\theta)}{G_p^2(\theta)} \frac{1+w(\theta)}{1+w_\Omega} - 1 \right] + \frac{1}{G_p(\theta)} \frac{1+w_\Omega}{1+w(\theta)} - 1 \right\}. \tag{3.13}$$

Nell'ipotesi di deboli correlazioni, presentandosi sempre sommate all'unità, la $w(\theta)$ e la w_Ω possono considerarsi trascurabili e l'ultima equazione può dunque ritenersi equivalente alla 3.10.

3.1.3 Conteggio delle coppie $\langle DR \rangle$

Nel seguito si calcola il valore atteso e la varianza del conteggio di coppie formate ad incrocio tra catalogo dato e catalogo random (DR). Considerando che i punti del catalogo dato e quello random non sono correlati, il valore atteso delle coppie risulta essere

$$\langle DR \rangle = n N_R G_p(\theta) \tag{3.14}$$

Il valore atteso dal suo quadrato risulta essere:

$$\langle DR \cdot DR \rangle = \left\langle \sum_{i \neq j}^{K} v_i p_j \Theta_{ij}^{\theta} \sum_{k \neq l}^{K} v_k p_l \Theta_{kl}^{\theta} \right\rangle,$$

dove p_l è il valore atteso della cella l contente un punto del catalogo random.

Il procedimento è del tutto analogo a quello effettuato per il conteggio di coppie di tipo DD, con la differenza che quando gli indici di due punti random si sovrappongono, la probabilità congiunta che ciò si verifichi è data da $(\frac{N_R}{K})^2$ anziché $\frac{N_R}{K}$ poiché si calcola su una media di cataloghi random per ogni data set. Dal calcolo si ricava:

$$\langle DR \cdot DR \rangle = n N_R^2 [G_p^2(\theta)(n-1) + G_t(\theta)]$$

e riprendendo la seconda equazione della 3.2, unita alla 3.14:

$$\langle \beta^2 \rangle = \frac{1}{n} \left[\frac{G_t(\theta)}{G_p^2(\theta)} - 1 \right].$$

3.1.4 Correlazione tra coppie di tipo DD e DR

L'ultimo termine da valutare è il valore atteso del prodotto di coppie di tipo DD e DR, che ha la seguente struttura:

$$\langle DD \cdot DR \rangle = \left\langle \sum_{i<j}^{K} v_i v_j \Theta_{ij}^{\theta} \sum_{k \neq l}^{K} v_k p_l \Theta_{kl}^{\theta} \right\rangle.$$

La somma può essere divisa in due termini: uno relativo alle configurazioni in cui gli indici dei tre punti del catalogo dato sono differenti tra loro ed un altro in cui due tre indici si sovrappongono. L'assenza della doppia degenerazione è data dall'impossibilità di sovrapposizione tra i punti del catalogo random ed i punti del catalogo dato. Per cui si ricava:

$$\langle DD \cdot DR \rangle = \frac{n(n-1)}{2} N_R [(n-2) G_p^2(\theta) + 2 G_t(\theta)].$$

Da questa, tenendo conto della terza equazione della 3.2, della 3.5 e della 3.14, segue:

$$\langle \alpha\beta \rangle = \frac{2}{n} \left[\frac{G_t(\theta)}{G_p^2(\theta)} - 1 \right].$$

3.2 Calcolo delle varianze e del bias degli stimatori proposti

Prima di procedere al calcolo delle varianze e del bias, è conveniente, p introdurre le seguenti quantità:

$$t = \frac{1}{n}\left[\frac{G_t(\theta)}{G_p^2(\theta)} - 1\right]$$

$$p = \frac{2}{n(n-1)}\left[\frac{1}{G_p(\theta)} - 2\frac{G_t(\theta)}{G_p^2} + 1\right] \simeq \frac{2}{n(n-1)G_p(\theta)}.$$

Conviene, inoltre, normalizzare i conteggi DD e DR ed i loro rispettivi valori attesi:

$$\begin{aligned} d =& \frac{2DD}{G_p(\theta)n(n-1)}, \ \Rightarrow \ \langle d \rangle = \frac{1+w(\theta)}{1+w_\Omega}, \\ x =& \frac{DR}{G_p(\theta)nN_R}, \qquad \Rightarrow \ \langle x \rangle = 1. \end{aligned} \tag{3.15}$$

Avendo introdotto tale notazione, si possono scrivere gli stimatori in una forma particolarmente semplice:

$$\begin{aligned} \frac{DD}{RR}: &\qquad 1 + w_{PH} = d \\ \frac{DD}{DR}: &\qquad 1 + w_{DP} = \frac{d}{x} \\ \frac{DD}{DR^2}: &\qquad 1 + w_H = \frac{d}{x^2} \\ \frac{DD - 2DR + RR}{RR}: &\qquad 1 + w_{LS} = d - 2x + 2. \end{aligned} \tag{3.16}$$

Calcolandone i relativi valori attesi e varianze, si ottiene:

$$\begin{aligned} 1 + \langle w_{PH} \rangle &= \langle d \rangle & var[w_{PH}] &\simeq \langle d \rangle (4t + p) \\ 1 + \langle w_{DP} \rangle &= \langle d \rangle (1 - t) & var[w_{DP}] &\simeq \langle d \rangle^2 (t + p) \\ 1 + \langle w_{H} \rangle &= \langle d \rangle (1 - t) & var[w_{H}] &\simeq \langle d \rangle^2 (p) \\ 1 + \langle w_{LS} \rangle &= \langle d \rangle & var[w_{LS}] &\simeq \langle d \rangle^2 (p). \end{aligned} \tag{3.17}$$

3.3 Conclusioni

Dall'equazione 3.17 e 3.15 si può notare come, nel limite di bassa correlazione, la varianza degli ultimi due stimatori sia molto vicina all'inverso del conteggio delle coppie, come ci si aspetterebbe da una distribuzione Poissoniana. In questi, infatti, il contributo alla varianza di $w(\theta)$ dato dal fattore $\langle DD \cdot DD \rangle$ si semplifica col quadrato del conteggio delle coppie. Il termine restante cresce come n^3 ed è compensato dal fattore DR; in questo modo per la varianza si ottiene un andamento Poissoniano che fornisce un'indeterminazione minore rispetto a quella valutata con i primi due stimatori.
É possibile affermare, dunque, che nel modello cosmologico proposto in cui le galassie si distribuiscono con andamento Poissoniano, il quarto stimatore dell'equazione 3.17, proposto da Landy e Szalay, risulta essere il più indicato per la stima della correlazione a due punti, permettendo infatti di minimizzare sia il bias che la varianza.

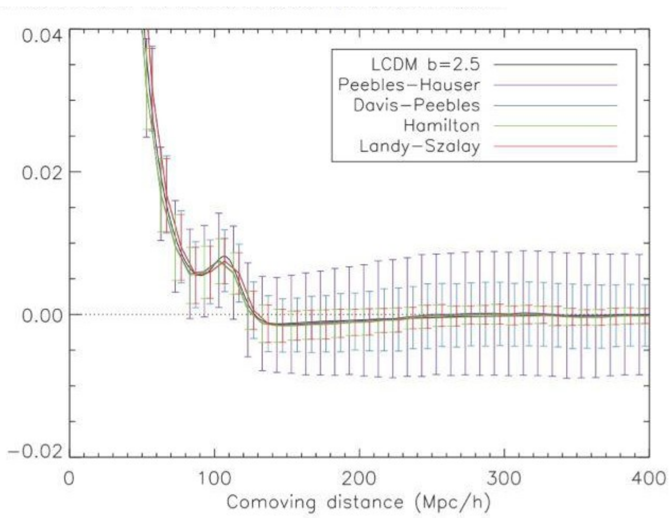

Figura 3.1:

Si mostra l'andamento della funzione di correlazione a due punti in funzione del raggio, con le varianze relative ai quattro estimatori esaminati. É possibile notare come quest'ultima sia minimizzata per lo stimatore di Landy&Szalay.

Bibliografia

[1] Peter Coles and Francesco Lucchin. *Cosmology: The origin and evolution of cosmic structure*. John Wiley & Sons, 2003.

[2] Stephen D Landy and Alexander S Szalay. Bias and variance of angular correlation functions. *The Astrophysical Journal*, 412:64–71, 1993.

[3] Malcolm S Longair. *Galaxy formation*. Springer Science & Business Media, 2007.

[4] Houjun Mo, Frank Van den Bosch, and Simon White. *Galaxy formation and evolution*. Cambridge University Press, 2010.

[5] Peter Schneider. Introduction and overview. In *Extragalactic Astronomy and Cosmology*, pages 1–43. Springer, 2015.

[6] Antoine Labatie, Jean-Luc Starck, Marc Lachièze-Rey, and Pablo Arnalte-Mur. Uncertainty in 2-point correlation function estimators and baryon acoustic oscillation detection in galaxy surveys. *Statistical Methodology*, 9 (1-2):85–100, 2012.

[7] Jean-Luc Starck, Antoine Labatie, and Marc Lachièze-Rey. Statistical study of the galaxy distribution. In Adnan Ghribi, editor, *Advances in Modern Cosmology*, chapter 8. InTech, Rijeka, 2011. doi: 10.5772/23010. URL https://doi.org/10.5772/23010.

[8] Y Wang, RJ Brunner, and JC Dolence. The sdss galaxy angular two-point correlation function. *Monthly Notices of the Royal Astronomical Society*, 432(3):1961–1979, 2013.

[9] Martin Kerscher, Istvan Szapudi, and Alexander S Szalay. A comparison of estimators for the two-point correlation function. *The Astrophysical Journal Letters*, 535(1):L13, 2000.