Random Gaussian Fields

They are ensembles of fields characterized by the fact that on making a realization (picking one field at random from the ensemble) the phases of the Fourier component, $\delta_{\vec{k}}$, are independent random variables for any two different values of $\vec{k}$. The ensemble is uniquely defined by the power spectra, $P(\vec{k})$, defined as the square of the modulus of the Fourier transform of the field.

For an infinite field, the power spectra is the same for every realization, with different realizations differing only in the phases of the Fourier transform, $\delta_{\vec{k}}$. However, when making a realization of a finite volume (using Fourier series rather than Fourier antitransform to represent the field) the value of $|\delta_{\vec{k}}|^2$ must change from one realization to another following a Rayleigh distribution determined by $P(k)$ so that:

$$\langle |\delta_{\vec{k}}|^2 \rangle = P(\vec{k})$$

This is strictly necesary if the realizations are to correspond to a random Gaussian field, but in practice the statistics of any observable in the field depending on many values of $\vec{k}$ would be the same if in all realizations we set $|\delta_{\vec{k}}| = P(\vec{k})^{1/2}$.

It is interesting to point out that a R.G.F with continuous $P(\vec{k})$ (including most physically interesting ones) is ergodic, in the sense that the statistics within N realizations of a box of size V are equall to the statistics of N boxes of volume V (sufficiently separated to avoid correlations) within a single realization of a much larger box.

At a randomly chosen point the value of a RGF is normaly distributed and the joint distribution of the fields at a randomly chosen point and at sets of positions determined by this point and certain displacement vectors is a Gaussian multivariate:

$$P(f_1 \cdots f_n) = \frac{e^{-\frac{1}{2}Q}}{[2\pi \,\mathrm{Det}\, C]^{1/2}} \quad ; \quad Q = \sum_{i,j} f_i f_j \left(C^{-1}\right)_{ij}$$

where C^{-1} is the inverse matrix of C

$$C_{ij} = \langle f_i f_j \rangle$$

The f's has been asumed to be central variables (i.e $\langle f \rangle = 0$) for non-central variables x we have: $f \equiv x - \langle x \rangle$ with $\langle f \rangle = 0$

Any field obtained from a RGF through a linear operation is also a RGF. For example if a RGF is filtered to remove fluctuations in scales smaller than r another RGF is obtained.

A joint distribution that plays an important role in theoretical computations for the L.S.S. is that for the field filtered on scale r_1 at a randomly chosen point, f_1, and the field filtered on scale r_2 at a distance R from the first, f_2. This can be useful, for example, to obtain the correlation at distance R between a cluster of galaxies (let's say with scale r_1) and galaxies (with scale r_2). The cosmological principle implies that the RGF of density fluctuations must be statistically isotropic $\Rightarrow P(\vec{K}) = P(K)$, and the correlation depends only on R and not in the direction of the vector from 1 to 2.

For what has been said the joint distribution of f_1, f_2 must be a Gaussian bivariate and all we need to determine is $\langle f_1^2\rangle, \langle f_2^2\rangle, \langle f_1 f_2\rangle$. But Parseval theorem tell us that

$$\langle f_1^2\rangle = \frac{1}{(2\pi)^3}\iiint |f_{1\vec{K}}|^2 d^3\vec{K} \overset{\text{for isotropy}}{=} \frac{4\pi}{(2\pi)^3}\int_0^\infty |f_{1K}|^2 K^2 dK$$

$$\langle f_1 f_2\rangle = \frac{1}{(2\pi)^3}\iiint f_{1\vec{K}} f^*_{2\vec{K}}\, d^3\vec{K}$$

Now we only need to express $|f_{1K}|^2, |f_{2K}|^2, f_{1K}f^*_{2K}$ in terms of $P(K)$. But f_1 is the convolution of the field δ with a function which is equal to $\frac{1}{V}$ inside a sphere of radius r_1 ($V \equiv \frac{4}{3}\pi r_1^3$) and zero outside and the Fourier transform of this function, WT_{r_1} is given by

$$WT_{r_1}(K) = 3\,\frac{\sin(Kr_1) - Kr_1\cos(Kr_1)}{(Kr_1)^3} \Rightarrow$$

$f_{1K} = \delta_K WT_{r_1}(K)$ and similarly for f_{2K}, but since the field f_2 is displaced a distance R from f_1, the Fourier transform of f_2 must have the aditional factor $e^{i\vec{K}\cdot\vec{R}}$. In computing $\langle f_2^2\rangle$ this factor does not appears, because it is $|f_{2K}|^2$ that enters, but for $f_{1K}f^*_{2K}$ we have:

$$f_{1K}f^*_{2K} = \delta_K WT_{r_1}(K)\left(\delta_K WT_{r_2}(K)e^{i\vec{K}\cdot\vec{R}}\right)^* = P(K)WT_{r_1}(K)WT_{r_2}(K)e^{-i\vec{K}\cdot\vec{R}}$$

Note that although $\langle f_1 f_2\rangle$ depends only on R on making the computation we must asume it to have some definite direction. Let's asume that it is in the z direction and use spherical coordinate with this direction as polar axis, then we may write:

$$\langle f_1 f_2\rangle = \frac{1}{(2\pi)^3}\int_0^\infty\int_0^{2\pi}\int_0^\pi P(K)WT_1WT_2\,e^{-iKR\cos\theta}\,\text{sen}\,\theta\, d\theta\, d\phi\, K^2 dK =$$

$$= \frac{1}{(2\pi)^2}\int_0^\infty\int_{-1}^1 P(K)WT_1WT_2\,e^{-iKR\mu}\,d\mu\, K^2 dK = \frac{1}{2\pi^2}\int_0^\infty P(K)WT_1WT_2\frac{\text{sen}(KR)}{KR}K^2 dK$$

with $\mu = \cos\theta$. The integral in μ is inmedeate.

1

Teorema de Wiener-Khinchin ("Parseval")

En analisis de Fourier, el teorema de Parseval nos dice que: $\iiint |f(\vec{x})|^2\, d^3\vec{x} = \frac{1}{(2\pi)^3}\iiint |f_{\vec{k}}|^2\, d^3\vec{k}$

o, mas generalmente; $\iiint f(x)\, g^*(x) = \frac{1}{(2\pi)^3}\iiint f_{\vec{k}}\, g^*_{\vec{k}}\, d^3\vec{k}$

, donde g, f son funciones de $\vec{x}$ para las que existe su transformada de Fourier, $g_{\vec{k}}$, $f_{\vec{k}}$, respectivamente.

Este Teorema nos permite expresar los promedios cuadráticos o las correlaciones en función de sus transformadas de Fourier. En el caso que nos interesa los campos g, f se obtienen del campo de fluctuaciones de densidad, $\delta(\vec{x})$, mediante operaciones lineales, por los cual sus transformadas se pueden expresar como el producto de la del campo, $\delta_{\vec{k}}$, por una función ventana (caracteristica de la operación lineal aplicada). Por tanto, podremos calcular sus varianzas y correlaciones en términos de el espetro de potencia de δ (i.e. $|\delta_{\vec{k}}|^2 \equiv P(\vec{k})$). Hay que señalar que, aunque en el curso de las demostraciones usemos $f_{\vec{k}}$ y $g_{\vec{k}}$, en las expresiones finales para cualquier cantidad estadística solo pueden aparecer $P(\vec{k})$ (no las $f_{\vec{k}}, g_{\vec{k}}$). Esto ha de ser así porque tanto las propiedades estadísticas como $P(\vec{k})$ caracterizan al colectivo completo de campos, por lo que no pueden depender de los componentes azarosos que entran en una realización concreta (las fases, que sí entran en $\delta_{\vec{k}}$); solo pueden depender del componente estructural ($P(\vec{k})$).

EL problema con el teorema de Parseval en nuestro caso es que tratamos con campos que se extienden indefinidamente en el espacio, por lo que las integrales espaciales de arriba divergen. Para tratar este caso hay que considerar volumenes finitos, V, (cajas de lado L) y tomar el Límite $V \to \infty$.

Para V finito, en lugar de integrales de Fourier tenemos series (valores de $\vec{k}$ discretos)

$$\delta(\vec{x}) = \frac{1}{L^{3/2}} \sum_{m,n,l} \delta_{\vec{k}_{m,n,l}}\, e^{i\vec{k}_{m,n,l}\cdot\vec{x}} \quad , \quad \vec{k}_{m,n,l} \equiv \frac{2\pi}{L}(m,n,l) \qquad (1)$$

La suma se extiende a todos los valores enteros (neg y pos.) de m,n,l, excluyendo $m=n=l=0$, que corresponde a una componente cte que, por definición, no puede tener δ ($\langle\delta\rangle = 0$).

$$\delta_{\vec{k}} \equiv \frac{1}{L^{3/2}} \iiint_V \delta\, e^{-i\vec{k}\cdot\vec{x}}\, d^3\vec{x} \qquad (2)$$

2 T: W-K

En clase di una definición diferente de $\delta_{\vec K}$ (sin $L^{3/2}$ en (2) y con L^{-3} en (1)), pero esta es mas conveniente.

Para el promedio cuadrático de δ tenemos, por definición:

$$\langle\delta^2\rangle = \frac{1}{V}\iiint_V \delta^2\, d^3\vec{x} = \frac{1}{V}\iiint \delta\,\delta^*\, d^3\vec{x} \qquad V\equiv L^3$$

sustituyendo (1) y su complejo conjugado en δ, δ^* ⇒

$$\langle\delta^2\rangle = \frac{1}{V}\iiint \frac{1}{L^{3/2}}\Big(\sum_{\vec K}\delta_{\vec K}\, e^{i\vec K\cdot\vec x}\Big)\frac{1}{L^{3/2}}\Big(\sum_{\vec K'}\delta^*_{\vec K'}\, e^{-i\vec K'\cdot\vec x}\Big)\, d^3\vec{x}$$

donde por $\sum_{\vec K}$ se entiende el sumatorio a todos los posibles valores de $\vec K$ (discretos) indicados en (1). $\sum_{\vec K'}$ es la misma suma, pero cambiando el simbolo del indice mudo $\vec K$. ⇒

$$\langle\delta^2\rangle = \frac{1}{V^2}\iiint_V \sum_{\vec K,\vec K'}\delta_{\vec K}\,\delta^*_{\vec K'}\, e^{i(\vec K-\vec K')\cdot\vec x}\, d^3\vec{x}$$

pero, $\iiint_V e^{i(\vec K-\vec K')\cdot\vec x}\, d^3x = 0$, excepto para $\vec K=\vec K'$, que vale V. ⇒

$$\langle\delta^2\rangle = \frac{1}{V}\sum_{m,n,l}|\delta_{\vec K_{m,n,l}}|^2 \qquad (3)$$

Habiendo escrito como en (1) la suma sobre los $\vec K$'s.

Supongamos que para $K<K_0$ $|\delta_{\vec K}|^2=0$ (o despreciable) ⇒ toda la contribución ~~a~~ esta suma vendrá de valores de $\vec K$ (con $K>K_0$), que para un V suficientemente grande corresponderá a números grandes de m,n,l. En estas condiciones al cambiar m,n,l en una unidad $|\delta_K|^2$ cambiará poco (si $P(K)$ continuo) y esta es justamente la condición para que una integral pueda ser bien aproximada por una suma (aunque en este caso la suma es lo exacto). ⇒

$$\frac{1}{V}\sum_{m,n,l}|\delta_{\vec K_{m,n,l}}|^2 \simeq \iiint_{m\,n\,l}|\delta_{\vec K_{m,n,l}}|^2\, d\frac{m}{L}\, d\frac{n}{L}\, d\frac{l}{L} =$$

$$\frac{1}{(2\pi)^3}\iiint_K |\delta_{\vec K}|^2\, dK_x\, dK_y\, dK_z$$

En el Limite de $V\to\infty$ en el que $m,n,l\to\infty$ para los K's que realmente contribuyen ⇒ la aproximación tiende a la igualdad:

$$\langle\delta^2\rangle = \frac{1}{(2\pi)^3}\iiint |\delta_{\vec K}|^2\, d^3\vec K$$

con:

$$|\delta_{\vec K}|^2 = \lim_{V\to\infty}\frac{1}{V}\left(\iiint_V \delta\, e^{-i\vec K\cdot\vec x}\, d^3\vec x\right)^2 \equiv P(K) \qquad (4)$$

La T. W-K suposición de que podemos encontrar un k_0 por debajo del cual (i.e, para escalas mayores) $|\delta_K|^2$ sea muy pequeño, se satisface en nuestro caso. Debido al P.C., en escalas suficientemente grandes $|\delta_K|^2$ debe ser tan pequeño como se quiera (i.e $\lim_{k\to 0} P(k)=0$).

En (4) vemos que en el límite de $V\to\infty$ $|\delta_K^2|$ es exactamente $P(K)$, es decir el mismo en todas las realizaciones (solo cambian las fases). En cambio, para V finito $|\delta_{\vec{k}}|^2$ fluctúa de una realización a otra de acuerdo con la siguiente distribución (de Rayleigh):

$$P(|\delta_{\vec{k}}|) = 2\, e^{-\frac{|\delta_{\vec{k}}|^2}{P(\vec{k})}} \frac{|\delta_{\vec{k}}|}{P(\vec{k})} \quad ; \quad \langle|\delta_{\vec{k}}|^2\rangle = \int |\delta_{\vec{k}}|^2 P(|\delta_{\vec{k}}|)\, d|\delta_{\vec{k}}| = P(\vec{k})$$

(5)

Para hacer una realización de $\delta(\vec{x})$ en V, se usa (1) con un número finito de sumandos, dependiendo de la resolución que se quiera tener. Llamando λ_r a esta resolución $\Rightarrow$ $K_{max} \sim 2\pi/\lambda_r$, lo cual determina los máximos valores de m, n, l $(\sim L/\lambda_r)$. A L/λ_r se le suele denominar rango dinámico (en las simulaciones) pues es el cociente entre la escala máxima y la mínima de las que se puede obtener información. Este número viene dado por el número de partículas que se usan en la simulación, N, $(L/\lambda_r \sim N^{1/3}$, para simulaciones en 3 dimensiones). Actualmente tenemos simulaciones con $N =$ varias veces 10^9.

Para una simulación con N partículas $(\Rightarrow \lambda_r \sim L/N^{1/3})$ el nº de modos que hay que usar para las condiciones iniciales ($\delta(\vec{x})$, se entiende que a $t=t_{in}$) es $\sim N$. Para cada uno de estos N modos (valores de $\vec{k}$) hemos de realizar $|\delta_{\vec{k}}|$ y θ_K: La forma general de realización de una variable aleatoria es:

$$W_i = \int_{x_i}^{\infty} P(x)\, dx \qquad (6)$$

donde x_i es la i-esima realización de la variable x, y W_i la i-esima realización de una variable aleatoria uniformemente distribuida entre 0 y 1.

Para θ_K la realización es trivial, pues esta uniformemente distribuida entre 0 y 2π: $P(\theta_K) = 1/2\pi$ $\Rightarrow$

$$\theta_{\vec{k}_i} = W_i\, 2\pi$$

Para $|\delta_{\vec{k}}|$, usando (5) se tiene:

$$|\delta_{\vec{k}}|_i^2 = -P(K)\, Ln(W_i) \qquad (7)$$

4 T. W-K

Como se ve la definición de $\delta_{\vec{K}}$ dada en (2) permite una identificación mas directa entre $|\delta_{\vec{K}}|^2$ para V finito y P(K) (: $\langle|\delta_K|^2\rangle = P(K)$ en lugar de $\langle|\delta_K|^2\rangle = VP(K)$ que se tendría con la definición que di en clase). Veamos que las dimensiones son correctas:

De la expresión : $\langle\delta^2\rangle = \frac{1}{(2\pi)^3}\iiint P(\vec{K})\, d^3\vec{K}$ (Teorema de W-K)

vemos que, puesto que δ es adimensional, $P(\vec{K})$ ha de tener las dimensiones de K^{-3}, es decir, dimensiones de volumen. En la definición (2) vemos que la integral tiene dimensiones de volumen por lo que : $[\delta_{\vec{K}}] = L^{3/2}$ (aquí L es "Longitud", no el lado de la caja) $\Rightarrow [|\delta_{\vec{K}}|^2] = L^3$.

$\Rightarrow [P(\vec{K})] = [|\delta_K|^2]$.

Problema 1

Consideren un campo gaussiano unidimensional ($\Rightarrow [P(\vec{K})] = L$) con $P(K) = cte = 10$Mpc (ruido blanco) y un "volumen" finito de 100Mpc. En este caso (1) queda en la forma :

$$\delta(\vec{x}) = \frac{1}{100^{1/2}} \sum_{m=1}^{100} 2\,\mathrm{Real}\left(\delta_{\vec{K}}\, e^{iK\cdot x}\right) \quad (8) \qquad K = \frac{2\pi\, m}{100\,\mathrm{Mpc}}$$

donde se ha tenido en cuenta que, por ser δ real $\Rightarrow \delta_{\vec{K}} = \delta^*_{-\vec{K}}$ y se han tenido en cuenta 100 modos.

1º) Demostrar que (7) se sigue de (5) y (6). Hacer 10^3 realizaciones de $|\delta_{\vec{K}}|$ para $K = 2\pi/100$Mpc usando (7); hacer el histograma y comprobar que "coincide" con (5) con el P(K) correspondiente.

2º) Hacer una realización de $\delta(x)$ usando (8) (haciendo la realización de los 100 valores de δ_K). Escoger 10^3 puntos al azar dentro del intervalo ($x_i = w_i\, 100$ Mpc), evaluar $\delta(x)$ en esos puntos, hacer el histograma y comprobar que se ajusta a una gaussiana. Calcular el promedio de δ^2 sobre esos valores y comprobar que está próximo a el miembro izquierdo de (3) ($\langle\delta^2\rangle$, que ha de ser exactamente igual a dicho miembro, está definido por la integral sobre el intervalo, y no coincide exactamente con el $\langle\rangle$ de δ^2 sobre puntos tomados al azar, pero difiere poco si se toman muchos puntos).

3º) Hacer 10^3 realizaciones de $\delta(x)$ usando solo dos modos en (8) (m=1, 2) comprobar que el histograma de los 10^3 valores de δ en un punto fijo cualquiera se aproxima a una gausiana, pero no el histograma de 10^3 puntos tomados al azar en una realización (i.e. el campo no es ergódico)

4º) En el apartado (3º) se ha visto que el campo con solo 2 modos no és ergódico, (lo cual solo puede ocurrir si $P(K)$ no es contínuo), pero sí es gaussiano. Veamos ahora que si en lugar de generar los $|\delta_K|$ de estos dos modos usando Rayleigh les asignamos en todas las realizaciones el valor fijo (para cada modo): $|\delta_K| = P(K)^{1/2}$, el campo deja de ser gaussiano. En particular, la dist. de Prob. de δ en un punto fijo (sobre realizaciones) deja de ser una gaussiana. Cual sea el punto fijo que se tome, es irrelevante, pues mientras que las fases estén uniformemente dist. entre 0 y 2π, el campo generado será estadisticamente homogeneo (lo que implica, en particular, que la dist. de δ es la misma en todos los puntos, aunque, obviamente, los histogramas correspondientes a distintos puntos difieren entre sí en el mismo grado en que cada uno de ellos difiere de la gaussiana teorica que se obtendría con un "numero" ∞ de realizaciones). Esta propiedad sigue manteniendose aunque las fases de distintos modos no fuesen variables aleatórias independientes (campos no-gaussianos), ~~mientras~~ siempre que las fases de cualquier modo considerado aisladamente, estén uniformemente dist. entre 0 y 2π sobre el conjunto de las realizaciones.

5º) Cuando determinamos las estadísticas de la estructura a gran escala (hablo en términos de esta, aunque las consideraciones son generales; puramente matemáticas) lo hacemos dentro de un volumen finito y con un número finito de galaxias (el campo δ basicamente lo muestreamos por medio de las galaxias, aunque las lentes gravitatorias proporcionan otra opción). Debido a esto, las estadísticas obtenidas tendrán un cierto error con respecto a las obtenidas de un volumen ∞ con un número de galaxias arbitrariamente grande, que son las que caracterizan el modelo (determinado por el P(K) del modelo (ej.: el de CDM), que a su vez viene determinado por los procesos físicos relevantes y el P(K) de las cond. In.). Al error debido al nº finito de galaxias se le suele denominar error de muestreo y tiende a cero cuando este $n^o \to \infty$ para un volumen dado. Pero aunque el error de muestreo sea despreciable, la finitud del volumen implica un error que suele denominarse "varianza cósmica" (aunque este nombre solo estaría justificado en el caso en que el volumen sea el de todo el universo observable; en otro caso el nombre de "error de estimación por volumen finito" parece mas adecuado).

En esta cuestión se pide determinar este último error (de L finito) para la cantidad estadística $\langle \delta^2 \rangle$. Para ello se harán 100 realizaciones del campo (cada una de ellas como la hecha anteriormente). Pero no hará falta obtener explícitamente $\delta(x)$, ya que:

$$\langle \delta^2 \rangle = \frac{1}{L}\int \delta^2(x)\,dx = \frac{1}{L}\sum_{n=-N}^{N} |\delta_{K_n}|^2 \underset{\delta_{K_n}=\delta^*_{K_{-n}}}{=} \frac{2}{L}\sum_{n=1}^{N} |\delta_{K_n}|^2$$

Por tanto, para cada realización $\langle \delta^2 \rangle$ viene inmediatamente dado por la suma de la derecha. Lo que hay que obtener es la desviación típica de esta cantidad:

$$\sigma(\langle \delta^2 \rangle) = \left(\mathrm{Var}(\langle \delta^2 \rangle)\right)^{1/2} = \left(\langle \langle \delta^2 \rangle^2 \rangle_{real} - \left(\langle \langle \delta^2 \rangle \rangle_{real}\right)^2\right)^{1/2}$$

Noten que $\langle \langle \delta^2 \rangle \rangle_{real}$ es justamente el valor que se obtendría en todas las realizaciones si en ellas se tomase $|\delta_K| = P(K)^{1/2}$. Vemos, por tanto, que lo que les comenté sobre el hecho de que cuando se tienen muchos modos la mayoría de las estadísticas no se ven muy afectadas por cambiar Rayleigh por esta prescripción, tiene sus salvedades: con $|\delta_K| = P(K)^{1/2}$ se obtendría bien $\langle \delta^2 \rangle$ (nos daría su valor medio), pero no podríamos obtener su varianza (que está directamente relacionada con el error de estimación de P(K) de volumen finito.