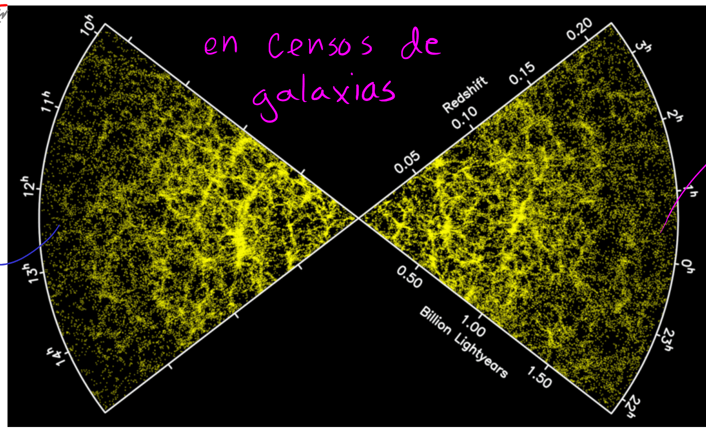
Función de Correlación de 2 puntos

en censos de galaxias



Clase 1

Buscamos patrones

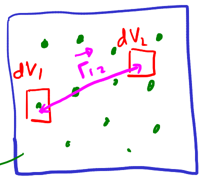
* Una manera de estudiarlos son las funciones de correlación de N-pts.

Estructura de Gran Escala del Universo

LSS - large scale structure

Para dist. aleatorias $dP_1 = \bar{\xi}\, dV_1$

N=2



densidad $\bar{\xi}$ promedio

Probabilidad de encontrar puntos en V_1 y V_2 es:

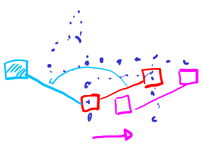
$$dP_{12} \equiv \bar{\xi}^2 \left(1 + \xi^{(2)}(\vec r_{12})\right) dV_1\, dV_2$$

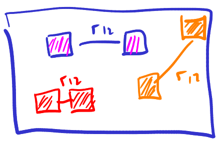
aleatorios

Define correlaciones por "arriba" de las dist. aleatorias

→ positivas son excesos

→ negativa son decrementos





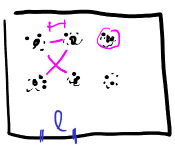
La función de correlación puede depender de la distancia entre volúmenes.

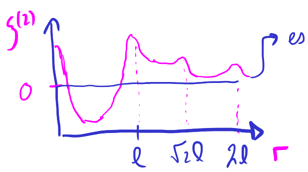
Si hay homogeneidad → (invarianza ante translaciones)

isotropía → (invarianza ante rotaciones)

(De forma estadística) ⇒ $\xi^{(2)}$ no depende de $\vec r_{12}$ sino sólo de su magnitud $|\vec r_{12}| \Rightarrow \xi^{(2)}(r)$

Ej.





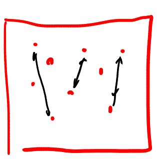
→ es como distribución aleatoria.

→ en 3 dimensiones sería $\sqrt{3}\,l$.

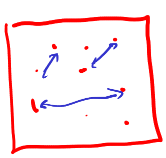
Pero ¿Cómo calculamos $\xi^{(2)}$?

Estimadores de $\xi^{(2)}$

* La idea es contar y en el límite de una infinidad de datos acercarnos al valor teórico de $\xi^{(2)}$
* Estamos midiendo distancias entre puntos de los datos.



D - Datos



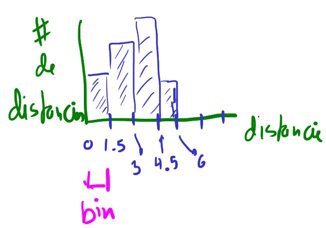
R - Aleatorios

Puedo calcular el número de distancias entre todos los puntos y las comparo con el mismo cálculo en la distribución aleatoria.

D: distancia 3.1, 3.2, 8.6, ...

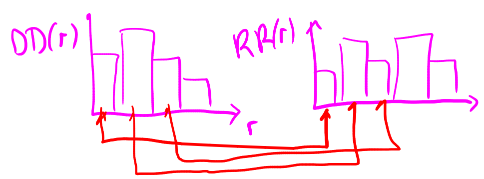
R = distancias: 3.3, 3.0, 8.5, ...

Mejor comparamos Histogramas



Al histograma de distancias en D le llamo DD

" " " " R " " RR



¿Cómo los comparo? Comparo para cada bin: ej bin no. 3

$$\frac{DD(r_3) - RR(r_3)}{RR(r_3)} = \frac{DD(r_3)}{RR(r_3)} - 1$$

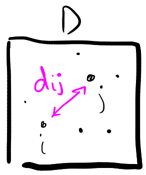
Es una normalización y tomamos RR porque son universales.

Estimador Preebles - Hauser $\quad \xi^{(2)}(r) = \dfrac{DD(r)}{RR(r)} - 1 = \dfrac{DD - RR}{RR}$

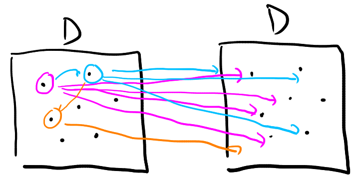
* Si el número de partículas es diferente entre D y R: Usamos histogramas normalizados $\dfrac{DD}{N_D}$, $\dfrac{RR}{N_R}$

escojo N_D para que $\sum_r DD(r) = 1$ y similar para N_R.

Algoritmo



es equivalente a



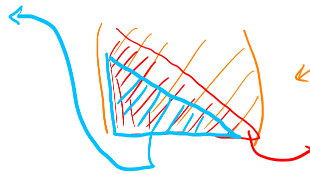
For $i=1$ to n
 For $j=1$ to n
  distancia[i][j] = ... pitágoras
  Histograma[bin de] += 1

1) Contamos doble ($d_{ij} = d_{ji}$)
 * Soln. For $j=i$ to n
  (ó 0 to i)

$N_D = \dfrac{n(n+1)}{2}$

2) contamos al mismo punto (d_{ii})
 * Soln: For $j=i+1$ to n
  (ó 0 to $i-1$).

$N_D = \dfrac{n(n-1)}{2}$



$N_D = n^2$

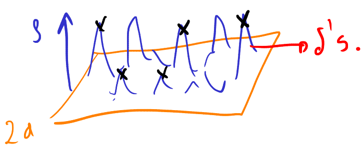
1).

∴ Para $n \gg 1$ estos códigos son $\Theta(n^2)$

Hay algoritmos más eficientes (basados en árboles, trees) Ej. KD-tree → disminuye a $\Theta(n \log n)$.

¿El estimador de PH es sesgado? i.e. difiere fuertemente de $\xi^{(2)}_{real}$ para muestras finitas?

Formalismo de Hamilton: Campo de densidad $\rho(\vec{x})$ y construyo "sobredensidades" $\delta(\vec{x}) = \dfrac{\rho(\vec{x}) - \bar{\rho}}{\bar{\rho}}$



ρ $\quad$ δ's. $\quad$ 2a

Tenemos galaxias: las escojo con una función de selección $W(\vec{x})$

ej galaxia 1 en $\vec{r}_1$ es $= \int_V \rho(\vec{x})\, W(\vec{x})\, d^3x$ donde $W(x) = \delta^3(\vec{x} - \vec{r}_1)$

V ↳ Volumen de los datos.

Definiciones: $\langle \, , \rangle \equiv \int_V d^3x \qquad \bar{\delta} \equiv \dfrac{\langle \delta(x) W(x) \rangle}{\langle W(x) \rangle}$

$\psi_n^{(k)}(\vec{r}_1, \ldots, \vec{r}_n) \equiv \dfrac{\langle W_1 \ldots W_n \delta_1 \ldots \delta_n \rangle}{\langle W_1 \ldots W_n \rangle} \quad (k \leq n)$

donde $W_i \equiv W(\vec{x}_i) \quad \delta_i \equiv \delta(\vec{x}_i)$

$= \dfrac{\int d^3\vec{x}_1\, d^3x_2 \ldots d^3x_n \boxed{W(\vec{x}_1) \ldots W(\vec{x}_n)}\, \delta(\vec{x}_1) \ldots \delta(\vec{x}_n)}{\int d^3\vec{x}_1 \ldots d^3\vec{x}_n\, W_1(\vec{x}_1) \ldots W(\vec{x}_n)}$

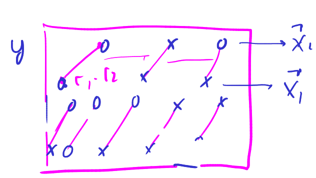
Noten:
* $\langle \delta \rangle = \langle \frac{\rho - \bar{\rho}}{\bar{\rho}} \rangle = \frac{\langle \rho \rangle - \bar{\rho}}{\bar{\rho}} = \frac{\bar{\rho} - \bar{\rho}}{\bar{\rho}} = 0$ (si $V \to \infty$)
* Número de galaxias $(n) = \langle \rho W \rangle = \int_V \rho(\vec{x})\, W(\vec{x})\, d^3x$
* $\bar{\delta} = \psi_1'$

↳ $W_1 \ldots W_n$ no siempre es separable ej $W_1 W_2 = W_1(\vec{x}_1) * W_2(\vec{x}_2)$

$W_1 W_2$ es la probabilidad conjunta de seleccionar 1 y 2.

Con este Formalismo estudiemos los histogramas DD y RR

$DD(\vec{r}_1, \vec{r}_2) = \langle \rho_1 \rho_2 W_1 W_2 \rangle \big|_{\vec{r}_1 - \vec{r}_2} = \int_{\vec{r}_1 - \vec{r}_2 = \vec{x}_1 - \vec{x}_2} \rho(\vec{x}_1)\, \rho(\vec{x}_2)\, W(\vec{x}_1) W(\vec{x}_2)\, d^3\vec{x}_1 d^3\vec{x}_2$



$\vec{r}_1 - \vec{r}_2$

$\delta = \dfrac{\rho - \bar{\rho}}{\bar{\rho}}$

$= \langle \bar{\rho}(1+\delta_1)\, \bar{\rho}(1+\delta_2)\, W_1 W_2 \rangle = \bar{\rho}^2 \big(\langle W_1 W_2 \rangle + \langle \delta_1 W_1 W_2 \rangle + \langle \delta_2 W_1 W_2 \rangle + \langle \delta_1 \delta_2 W_1 W_2 \rangle \big)$

$$= \bar{\rho}^{-2}\langle W_1 W_2\rangle\left(1+\frac{\langle\delta_1 W_1 W_1\rangle}{\langle W_1 W_2\rangle}+\frac{\langle\delta_2 W_1 W_2\rangle}{\langle W_1 W_2\rangle}+\frac{\langle\delta_1\delta_2 W_1 W_2\rangle}{\langle W_1 W_2\rangle}\right) = \underline{\bar{\rho}^{-2}\langle W_1 W_2\rangle}\left(1+\Psi_2'(\vec{r}_1)+\Psi_2'(\vec{r}_2) + \Psi_2^2(\vec{r}_1,\vec{r}_2)\right)$$

$$RR(r_1,r_2) = \bar{\rho}^{-2}\int W_1 W_2\, d^3\vec{x}_1\, d^3\vec{x}_2 = \underline{\langle W_1 W_2\rangle}\,\bar{\rho}^{-2}$$

es una corrección o sesgo?

$$\therefore\ \xi^{(2)}_{PH} = \frac{DD}{RR}-1 = \underbrace{\Psi_2^2(\vec{r}_1,\vec{r}_2)}_{\text{Esta es la teórica}} + \Psi_2'(r_1)+\Psi_2'(r_2) = \xi^{(2)}_{real} + 2\Psi_2'$$

para $n_D \neq n_R$ $\quad \xi^{(2)}_{PH} = \frac{n_R^2}{n_D^2}\frac{DD}{RR}-1$

$$n_D = \langle \rho W\rangle = \langle\bar{\rho}(1+\delta)W\rangle = \langle W\rangle + \langle\delta W\rangle = \underbrace{\langle W\rangle}_{n_R}\left(1+\frac{\langle\delta W\rangle}{\langle W\rangle}\right) = n_R(1+\bar{\delta})$$

$$\xi^{(2)}_{PH} = \frac{1}{(1+\bar{\delta})^2}\left(\xi^{(2)}_{real}+\Psi_2'(\vec{r}_1)+\Psi_2'(\vec{r}_2)\right) = \xi^{(2)}_{real} + \underline{\Psi_2'(\vec{r}_1)} + \underline{\Psi_2'(\vec{r}_2)} - 2\underline{\bar{\delta}} + \mathcal{O}(\bar{\delta}^2, \Psi\bar{\delta}, \bar{\delta}\xi)$$

$\bar{\delta} \ll 1$

Podemos tener correcciones menos relevantes:

Davis-Peebles $\quad \xi^{(2)}_{DP} = \frac{1}{N_{est}}\frac{DD}{DR}-1$

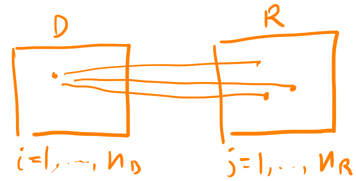
Hamilton $\quad \xi^{(2)}_{Ham} = \frac{DD\ RR}{DR}$

Landy-Szalay $\quad \xi^{(2)}_{LZ} = 1+\frac{1}{N_{est}^2}\frac{DD}{RR} - 2\frac{1}{N_{est}}\frac{DR}{RR}$

las correcciones son cuadráticas en Ψ_2' y $\bar{\delta}$.

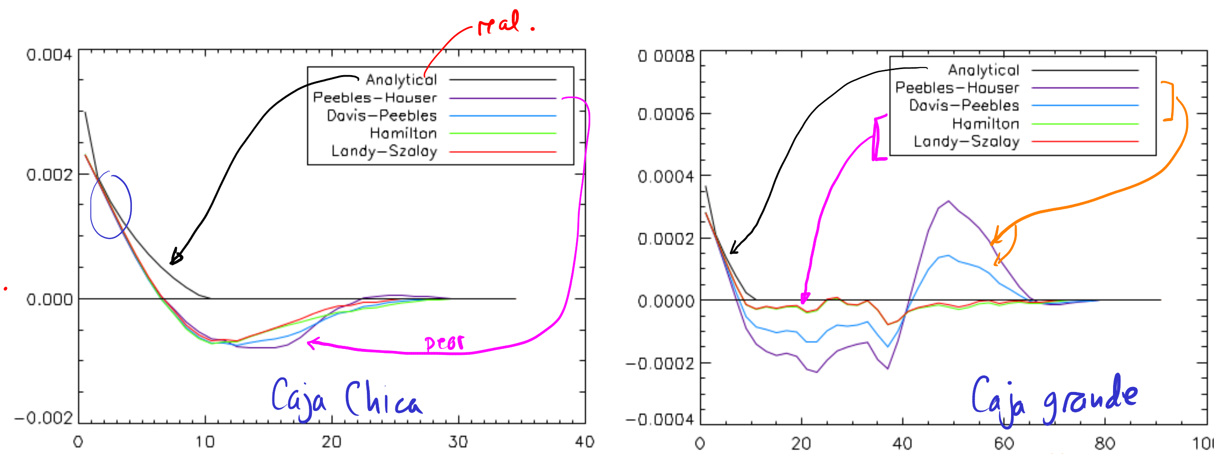
Donde $N_{est} = n_D/n_R$

$DR \to$ es un histograma que "mezcla" datos con aleatorios

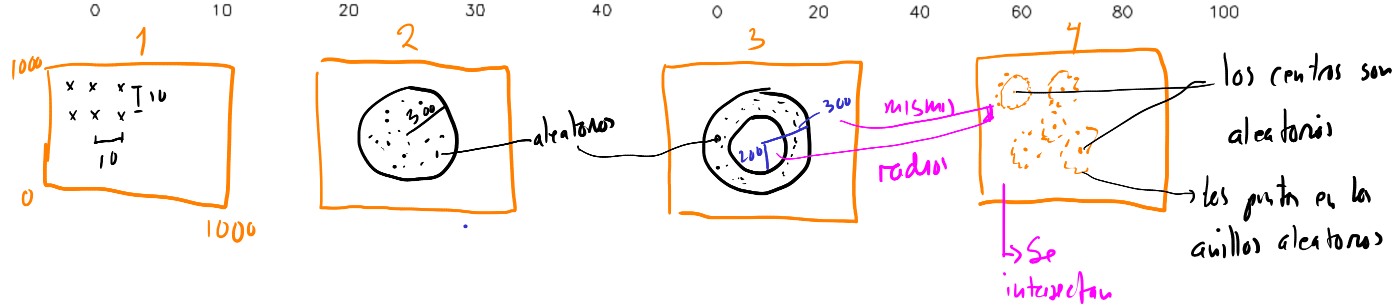


Ejemplo

Real se obtiene de un Proceso de Cox.



Cajas



Referencias

https://arxiv.org/abs/2006.05434
Hamilton: http://articles.adsabs.harvard.edu/pdf/1993ApJ...417...19H