

Proyecto Programación básica: Función de Correlación de dos puntos para una distribución de puntos en 3 dimensiones.

Prof. Alma González

November 26, 2018

Las funciones de correlación de dos puntos, son ampliamente usadas para analizar sistemas de muchos cuerpos, nos dan información sobre la estructura y evolución de sistemas, por ejemplo sistemas de n-cuerpos gravitatorios, o bien gases, líquidos o sólidos. También es de utilidad en otras áreas, incluso en economía. De forma muy práctica definiremos la función de correlación de dos puntos como una medida de que tanto se parece una distribución de puntos a una distribución aleatoria. Si la distribución que estamos analizando es aleatoria, la distribución de puntos será cero. Si la función de distribución tiene alguna estructura, por ejemplo los puntos están mas aglomerados en ciertas escalas, entonces será distinta de cero y tendrá formas muy particulares. Otra forma de entender la función de correlación es hacer la pregunta, ¿cuál es la probabilidad de tener un par de puntos separados a una cierta distancia r ?

La función de correlación de dos puntos, para una distribución de puntos la calcularemos usando dos estimadores, el usuario del programa deberá elegir que estimador usar.

1) Estimador simple:

$$\xi(r) = \frac{n_R}{n_D} \frac{DD}{RR} - 1 \quad (1)$$

2) Estimador de Landy-Szalay, principalmente usado en sondeos de galaxias.

$$\xi(r) = \frac{(DD/n_D - 2DR/n_{DR} - RR/n_R)}{RR/n_R} \quad (2)$$

Donde $r = |\vec{r}_2 - \vec{r}_1|$ es la distancia entre dos puntos; DD es el número de pares separados a una distancia r en el archivo de datos, DR es el número de pares separados a una distancia r para el archivo de datos y el archivo que contiene puntos distribuidos aleatoriamente, como si fueran un solo conjunto de datos, y RR es el número de pares separados a una distancia r para el archivo de puntos aleatorios. n_R y n_D son el numero de pares totales en el archivo de aleatorios y el archivo de datos. DD, DR, RR son histogramas, por lo que debemos definir un intervalo de distancias r , tamaño de bin, para construir dichos histogramas.

El proyecto consiste en:

1. Escribir un programa para calcular la función de correlación, el cuál debe leer un archivo que contiene las coordenadas (x,y,z) de los puntos con los cuales se va a calcular la función de correlación. También deberá tener la opción de leer otro archivo que contenga los datos de puntos distribuidos de forma aleatoria o bien de generar dicho archivo, en caso de que no exista. El archivo de puntos aleatorios debe tener la misma geometría que el archivo de datos. Por simplicidad será un cubo de lado L (definido por el usuario). El formato de los archivos en ambos casos consiste de 3 columnas (x,y,z), y N filas, donde N es el numero de datos. En general el numero de puntos en los archivos de datos y los aleatorios pueden ser diferentes.
2. Se debe calcular la función de correlación de acuerdo al estimador que elija el usuario.

3. Al terminar se debe guardar un archivo de dos columnas el radio y la función de correlación en dicho radio.
4. El programa debe hacer uso de funciones, de multiples archivos para la organización de todo el programa, tener un archivo de compilación (makefile), y usar la mayor cantidad de conceptos vistos en clase (ej. uso dinámico de la memoria)(4 puntos). Se obtendrá mejor puntaje cuanto mas organizado, y bien diseñado esté el programa.
5. Se deben presentar pruebas del funcionamiento del código que muestren que cuando se calcula la función de correlación de un archivo con datos distribuidos de forma aleatoria, la función de correlación es cero (1 punto).
6. Se les proveerán 3 archivos de datos diferentes, para los cuales se calculará la función de correlación y se guardaran en archivos diferentes, los cuales posteriormente se usarán para graficar $\xi(r)$ para cada archivo. El objetivo es que después de usar el código, se pueda ver a partir de la función de correlación que los tres archivos son diferentes. En los resultados se debe discutir, el efecto de usar diferentes números de bin para calcular la función de correlación así como el efecto de usar los dos estimadores diferentes. (3 puntos)
7. Para el reporte se debe investigar un poco sobre el uso de funciones de correlación de dos puntos diferentes áreas. (2 puntos)
8. Se recomienda hacer pruebas primero con archivos pequeños (creados por ustedes) para verificar el funcionamiento, y solo usar los 3 conjuntos de archivos que se proveerán para el proyecto cuando ya se tenga certeza que el programa funciona correctamente, pues el programa puede tardar bastante tiempo en ejecutarse.
9. Extra: suponiendo que esperamos que haya una distancia característica en la función de correlación y que ésta es menor a la distancia máxima entre los puntos de la distribución de los datos, ¿cómo podrías hacer que el código sea mas eficiente y se ejecute más rápido? Obtendrás puntos extra si lo implementas y demuestras que efectivamente funciona.