



Trabajo Practico 3: Astrometría. Base de datos

Camila Sofia Beatriz Hormaeche

¹ Facultad de Matematica, Astronomia, Fisica y Computacion (UNC)

Received: ... / Accepted: ...

©The Authors 2024

Resumen /

Keywords /

1. Introducción

1.1. Base de Datos

En general, la información de grandes relevamientos, simulaciones, etc., se encuentra almacenada en bases de datos. Uno de los paradigmas más utilizados para organizar bases de datos es el de **Bases de Datos Relacionales**. En este tipo de modelo, la información se organiza en tablas, las cuales se relacionan entre sí a partir de alguna propiedad de los datos. Cada tabla es un conjunto de registros.

Existen programas o sistemas de gestión de bases de datos relacionales. Entre los más conocidos, se destacan, por ejemplo, MySQL, PostgreSQL, Oracle y Microsoft SQL Server. Cabe mencionar que recientemente se desarrolló el software **SciDB**, el cual está preparado para manipular información científica (de hecho, fue desarrollado para el LSST).

Uno de los lenguajes más utilizados para manipular y acceder a bases de datos es el denominado **SQL** (Structured Query Language). Con este lenguaje se pueden crear, modificar y consultar bases de datos. SQL se divide en dos partes: un lenguaje de manipulación de datos y un lenguaje de definición de datos.

El objetivo de este trabajo es lograr una familiarización con el manejo de bases de datos, pudiendo extraer información, procesarla y analizarla mediante diferentes métodos y principios de probabilidad y estadística. Se propone utilizar, en particular, la base de datos del **Sloan Digital Sky Server** para extraer datos de galaxias y estudiar su morfología, corrimiento al rojo, magnitudes, entre otras características.

1.2. Procedimiento

Primero, ingresamos al sitio (?), CasJob, creamos una cuenta de usuario y analizamos el sitio para familiarizarnos con las tablas y datos que proporciona. Se exploró la estructura de la base de datos usando **Sky-server** → **SQL-Search** para la obtención de la tabla de datos. A esta tabla se le pidió que contuviera datos de galaxias, en particular magnitudes petrosianas en las bandas u, g y r, redshift y clasificación como elíptica o

espiral.

El query que se utilizó contiene lo siguiente:

```
SELECT
  s.specObjID, g.spiral, g.elliptical, s.petroMag_u, s.petroMag_r, s.petroMag_g
INTO mydb.MyTable_1
FROM SpecPhotoAll AS s
JOIN zooSpec AS g ON g.specobjid = s.specObjID
WHERE
  ((g.spiral > 0.8) OR (g.elliptical > 0.8))
```

Notar que se descargaron todos los datos disponibles teniendo en cuenta este query. A continuación, se descargo la tabla de datos para resolver una serie de ejercicios.

2. Ejercicio 2

Estudie la distribución de índices de color g-r y u-g para galaxias elípticas y espirales. Determine si, para ambos casos, las distribuciones son consistentes. Discuta la validez de la forma de la distribución propuesta para el estadístico de la prueba.

Para resolver esto, primero se graficó la distribución de galaxias elípticas y galaxias espirales. De esta manera, se puede observar la distribución dependiendo del color (g o r), como se observa en la Figura 1:

Además, se utilizó el método de Kolmogorov-Smirnov para evaluar si las distribuciones son diferentes, haciendo además una prueba de hipótesis, como se observa en la Figura 2:

El método de Kolmogorov-Smirnov es una prueba estadística no paramétrica que se utiliza para comparar dos distribuciones de probabilidad. Esta prueba se basa en la diferencia máxima entre las funciones de distribución acumulativa (CDF) de dos muestras. El estadístico de Kolmogorov-Smirnov se define como:

$$D = \max |F_n(x) - G_m(x)|$$

donde $F_n(x)$ y $G_m(x)$ son las CDFs de las dos muestras. Un valor de D alto indica que las distribuciones son diferentes. Para evaluar la significancia de esta diferencia, se compara el estadístico con valores críticos obtenidos de tablas o se calcula un valor p . Si el valor p es menor que el nivel de significancia seleccionado

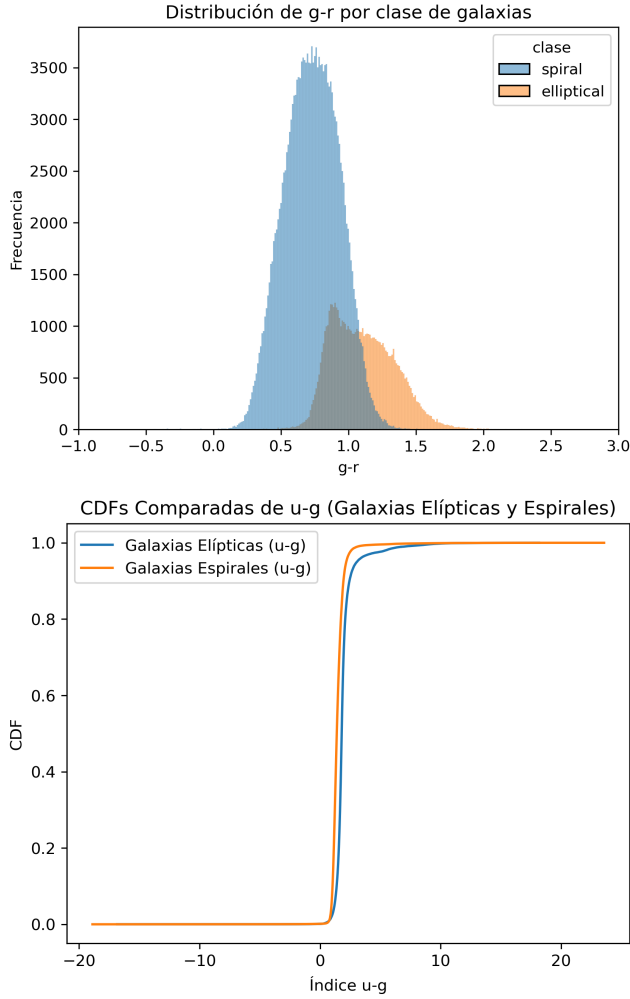


Fig. 1. Comparación entre la distribución por clase de galaxias color-color

(usualmente 0.05), se rechaza la hipótesis nula de que las dos muestras provienen de la misma distribución.

Se obtuvo que ambas distribuciones, difieren.

3. Ejercicio 3

Se busca estudiar una distribución de tipos morfológicos y determinar si es consistente con una distribución uniforme. Se puede utilizar una prueba estadística que compare la distribución observada con una distribución uniforme teórica. En el contexto de este ejercicio, se supone que el número de galaxias elípticas y espirales debería ser el mismo si la distribución es uniforme.

Para comprobar esto, se realiza una prueba de χ^2 , la cual se utiliza para evaluar si las frecuencias observadas en diferentes categorías son consistentes con las frecuencias esperadas bajo una hipótesis específica (en este caso, la hipótesis de que las galaxias están uniformemente distribuidas).

En la figura 3, se grafica la distribución de galaxias elípticas y espirales, teniendo en cuenta la distribución esperada y la real; se observa que las distribuciones no coinciden.

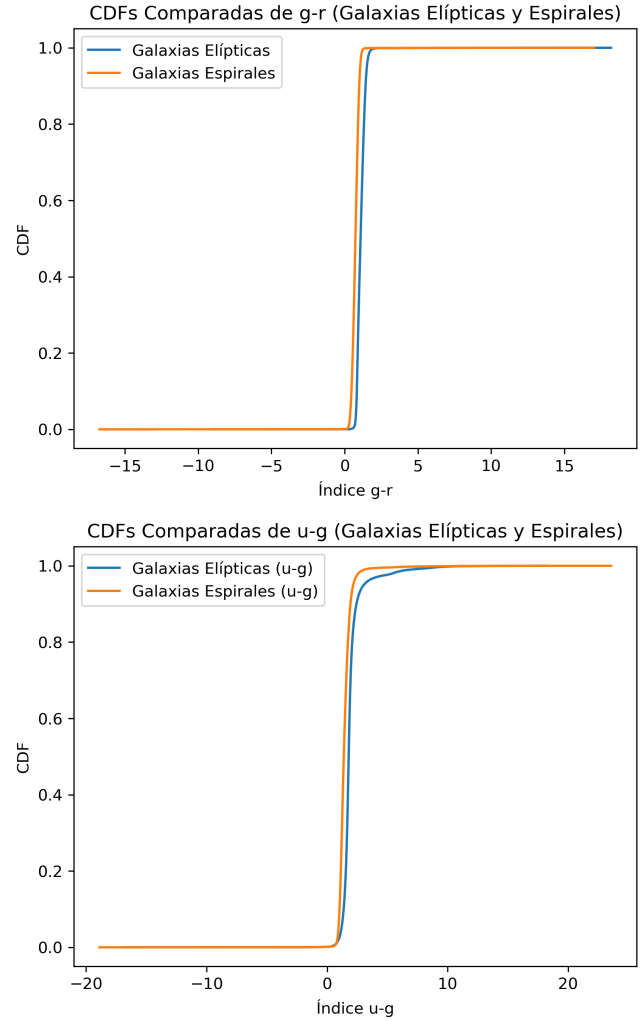


Fig. 2. Método de Kolmogorov-Smirnov

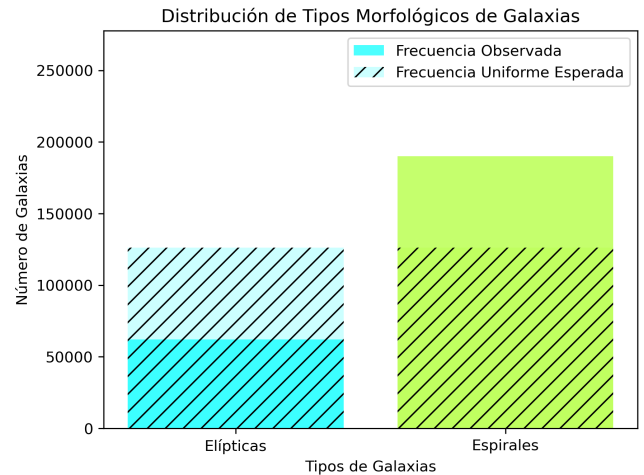


Fig. 3. Se grafica la distribución de galaxias elípticas y espirales, teniendo en cuenta la distribución esperada y la real.

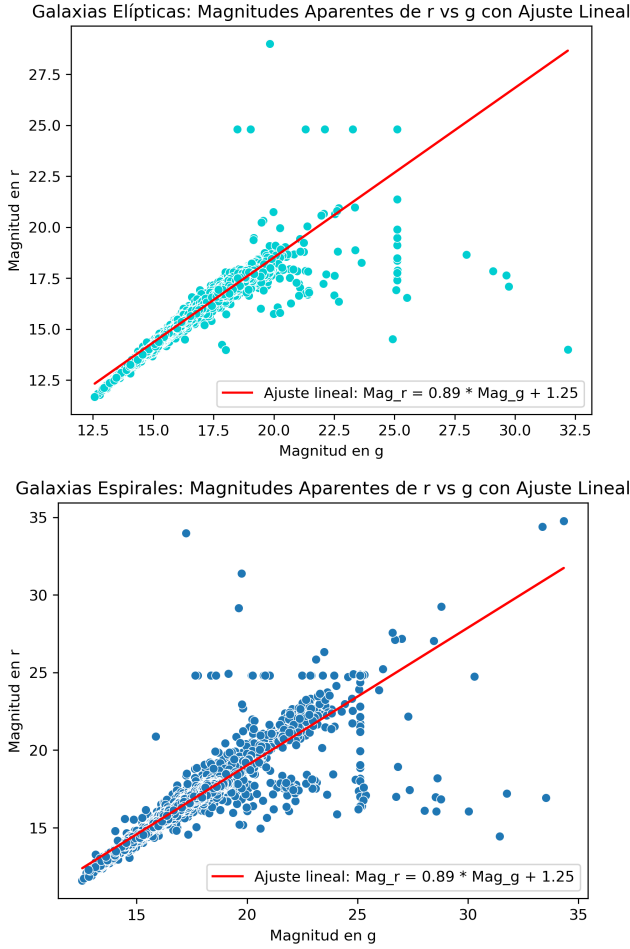


Fig. 4. Ajuste lineal entre magnitudes aparentes r vs g.

4. Ejercicio 4

Grafique las magnitudes aparentes de galaxias en la banda r en función de las magnitudes en la banda g, y obtenga un ajuste para la relación entre ambas. La magnitud aparente es una medida del brillo de un objeto astronómico, tal como lo percibe un observador en la Tierra. El ajuste lineal se utiliza para modelar la relación entre dos variables. En este caso, queremos ajustar una línea recta a los datos de magnitudes en las bandas g y r.

A través de gráficos, como el que se muestra en la figura 4, de dispersión y ajustes lineales, se puede observar cómo se relacionan estas magnitudes para diferentes tipos de galaxias (elípticas y espirales). Los resultados del ajuste (pendiente e intersección) proporcionan información adicional sobre la relación entre estas magnitudes.

5. Ejercicio 5

Calcule la magnitud absoluta para cada galaxia, usando la aproximación:

$$M = m - 25 - 5 \cdot \log_{10} \left(\frac{c \cdot z}{H} \right)$$

donde c es la velocidad de la luz y $H =$

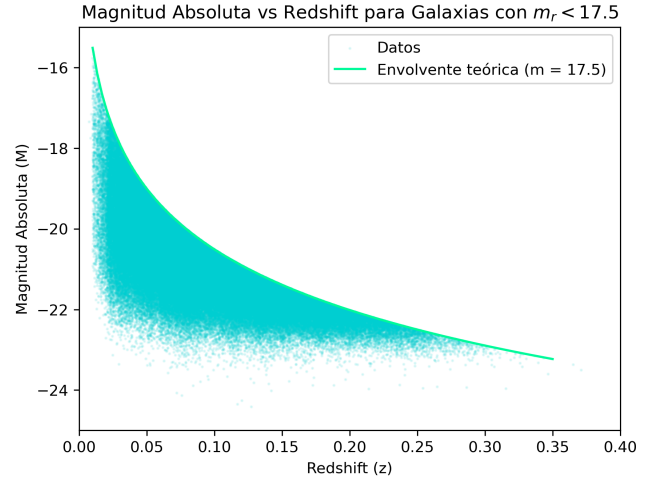


Fig. 5. Se grafica la magnitud absoluta en función del redshift, observando la tendencia de los datos.

75 km s⁻¹/Mpc. Grafique la magnitud absoluta vs. el redshift para todas las galaxias con $m_r < 17.5$ y obtenga un ajuste para la envolvente de los puntos. Discuta en el informe el origen de la forma de los puntos y el procedimiento para el ajuste del modelo. La magnitud absoluta es una medida del brillo intrínseco de un objeto astronómico, corregido por la distancia.

El redshift es un desplazamiento hacia el rojo de la luz proveniente de objetos astronómicos, que se relaciona con la expansión del universo. Cuanto mayor es el redshift, más distante y, en general, más antiguo es el objeto.

Al graficar la magnitud absoluta en función del redshift, se pueden observar patrones que se pueden modelar. La envolvente es una línea que representa la tendencia general de los puntos y se utiliza para entender mejor la relación entre las dos variables. Para obtener la envolvente teórica, se calcula la magnitud absoluta para un rango de redshifts utilizando la fórmula dada. Esto proporciona un modelo que representa la tendencia general de los datos.

En la figura 5, se observa la tendencia de los datos y se aprecia un corte. Esto se debe a que el catálogo utilizado asegura que, a magnitudes aparentes $r < 17.5$, se puede garantizar que lo que se observa son galaxias. Además, al aumentar la distancia, las galaxias más débiles no se observarán.

References