

CLP Lab 0 Report

Albert Aparicio Isarn
albert.aparicio.isarn@alu-etsetb.upc.edu

Héctor Esteban
hect.esteban@gmail.com

1. Test de gaussianidad de distribuciones sintéticas

Como se puede ver en la figura 1, los gráficos de las distribuciones *Rayleigh*, *Laplaciana* y *Uniforme* son visiblemente distintos a los de la distribución.

Estas diferencias se ven más claramente en los *NormPlot* y los gráficos de *Skewness* y *Kurtosis*.

2. Observación de base de datos Brain

Para los gráficos de esta sección se ha elegido mostrar las características de la clase **1 - Materia Gris**.

2.1. Observe y comente todas las gráficas obtenidas

En la figura 2 se pueden ver los gráficos para materia gris.

En la figura 2b se puede ver que las características 2 y 5 son las que tienen un *Skewness* y *Kurtosis* más parecidos a una distribución gaussiana

En el *Scatter plot* de la figura 2c se ve como los pares de características 2 o 3 con 4 o 5 no son particularmente discriminativos.

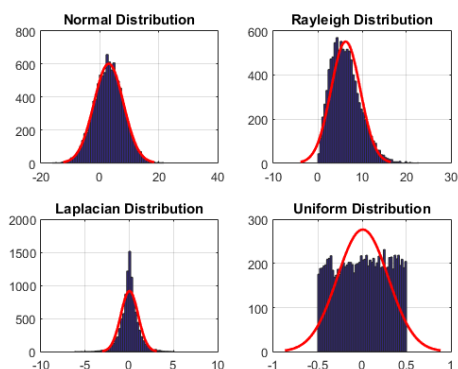
En la figura 2d vemos como el gráfico de CDF de la característica 1 es el que más se acerca al de una distribución gaussiana.

Los histogramas de la figura 2e vemos como la característica 4 es la más parecida a una distribución gaussiana, mientras que el resto de ellas muestran histogramas "picudos."o asimétricos.

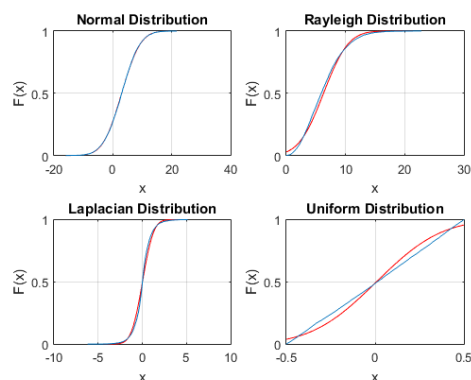
Según los *NormPlot* de la figura 2f, la característica 2 está más cercana a ser gaussiana, ya que el resto muestran severas diferencias respecto la recta de referencia.

2.2. A partir de la observación del "scatter plot", proponga el subconjunto de 2 características o coordenadas que resulte más discriminativo para este caso

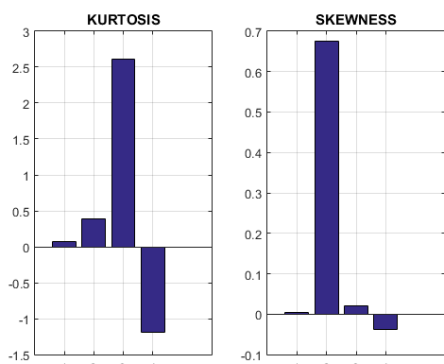
En el *Scatter Plot* de la figura 2c se ve como la característica 1 es la que más separabilidad proporciona, especialmente con la característica 4.



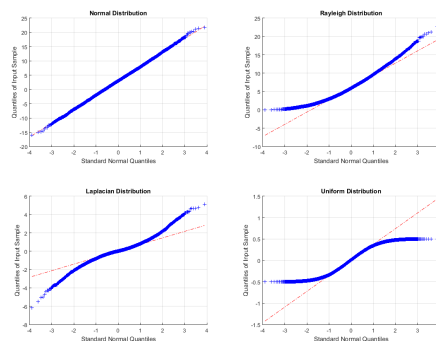
(a) Histogramas de las distribuciones.



(b) Funciones densidad de probabilidad de las distribuciones.



(c) Gráficos de *Skewness* y *Kurtosis* de las distribuciones.



(d) Gráficos *NormPlot* de las distribuciones.

Figura 1: Gráficos de los tests de gaussianidad de las cuatro distribuciones sintéticas.

Esta pareja de características concentra los puntos del gráfico de una forma más separada y compacta que el resto de combinaciones.

2.3. Cambie la clase analizada en la última parte del programa y copie los resultados obtenidos en el documento de resultados

Las figuras 3 y 4 muestran los conjuntos de gráficos obtenidos para las dos clases restantes, **2 - Materia blanca** y **3 - Líquido cefaloraquídeo**, respectivamente.

3. Estimación de intervalos de confianza

En la tabla 1 se muestran los intervalos de confianza calculados con el siguiente código, añadido al script `prac0_Brain.m`:

```
% Código usado en el ejercicio 2.4
% Intervalos clase 1
index=find(Labels==1);

alpha95 = 1 - 0.95;
alpha99 = 1 - 0.99;
alpha999 = 1 - 0.999;

x = mean(X(index,1));
s2 = var(X(index,1));
n = length(index);
t95 = tinv(1-(alpha95/2), n-1);
t99 = tinv(1-(alpha99/2), n-1);
t999 = tinv(1-(alpha999/2), n-1);

interval_p_1_95 = x + t95 * (sqrt(s2/n));
interval_n_1_95 = x - t95 * (sqrt(s2/n));
interval_p_1_99 = x + t99 * (sqrt(s2/n));
interval_n_1_99 = x - t99 * (sqrt(s2/n));
interval_p_1_999 = x + t999 * (sqrt(s2/n));
interval_n_1_999 = x - t999 * (sqrt(s2/n));

% Intervalos clase 2
index=find(Labels==2);

x = mean(X(index,1));
s2 = var(X(index,1));
n = length(index);
t95 = tinv(1-(alpha95/2), n-1);
t99 = tinv(1-(alpha99/2), n-1);
t999 = tinv(1-(alpha999/2), n-1);

interval_p_2_95 = x + t95 * (sqrt(s2/n));
interval_n_2_95 = x - t95 * (sqrt(s2/n));
interval_p_2_99 = x + t99 * (sqrt(s2/n));
interval_n_2_99 = x - t99 * (sqrt(s2/n));
interval_p_2_999 = x + t999 * (sqrt(s2/n));
interval_n_2_999 = x - t999 * (sqrt(s2/n));

% Intervalos clase 3
index=find(Labels==3);

x = mean(X(index,1));
s2 = var(X(index,1));
n = length(index);
t95 = tinv(1-(alpha95/2), n-1);
t99 = tinv(1-(alpha99/2), n-1);
t999 = tinv(1-(alpha999/2), n-1);
```

```

interval_p_3_95 = x + t95 * (sqrt(s2/n));
interval_n_3_95 = x - t95 * (sqrt(s2/n));
interval_p_3_99 = x + t99 * (sqrt(s2/n));
interval_n_3_99 = x - t99 * (sqrt(s2/n));
interval_p_3_999 = x + t999 * (sqrt(s2/n));
interval_n_3_999 = x - t999 * (sqrt(s2/n));
../codigo_24.m

```

En el código, las variables `interval_p_X_Y` se refieren al intervalo de confianza positivo para la clase X, con nivel de confianza Y, y las variables `interval_n_X_Y` se refieren al intervalo de confianza negativo para la clase X, con nivel de confianza Y.

Clase Confianza	Materia gris	Materia blanca	Líquido cefaloraquídeo
95 %	{0,0456, 0,0544}	{0,3778, 0,3844}	{-0,6320, -0,6225}
99 %	{0,0443, 0,0557}	{0,3767, 0,3854}	{-0,6335, -0,6210}
99.9 %	{0,0427, 0,0573}	{0,3755, 0,3866}	{-0,6352, -0,6192}

Cuadro 1: Intervalos de confianza para las tres clases de la BBDD y para los tres niveles de confianza requeridos.

4. Test de gaussianidad de base de datos Brain

A continuación se muestra el código que se ha añadido al script `prac0_Brain.m`:

```

% Codigo usado en el ejercicio 2.5
% Hipotesis clase 1
index=find(Labels==1);

[h1,p1,stats1] = chi2gof(X(index,1),'Alpha',0.001);
df1 = stats1.df;

% Hipotesis clase 2
index=find(Labels==2);

[h2,p2,stats2] = chi2gof(X(index,1),'Alpha',0.001);
df2 = stats2.df;

% Hipotesis clase 3
index=find(Labels==3);

[h3,p3,stats3] = chi2gof(X(index,1),'Alpha',0.001);
df3 = stats3.df;
../codigo_25.m

```

Las variables `hX`, `pX`, `dfX` contienen la hipótesis seleccionada, el valor del *p-value* y los grados de libertad para la clase X, respectivamente.

4.1. Hipótesis nula H_0 : La característica 1 de la clase 1 es gaussiana con un nivel de significancia (α) de 0.001

La función devuelve $h = 1$ (rechaza la hipótesis nula).

El valor de p -value es $2,22 \cdot 10^{-71} \approx 0$. Los grados de libertad son 4.

Si miramos la Kurtosis y el Skewness de la clase Materia Gris en la primera característica (figura 2b, 1ra columna) se ve con claridad que no corresponden a una Gaussiana.

4.2. Hipótesis nula H_0 : La característica 1 de la clase 2 es gaussiana con un nivel de significancia (α) de 0.001

Se mantendrá la hipótesis nula $h = 0$.

El valor de p es $p - value = 0,0012 > \alpha$. Los grados de libertad son 7.

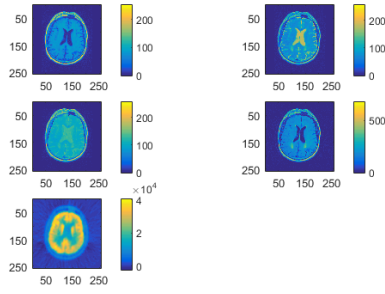
Si miramos el Skewness y Kurtosis (figura 2b, 2da columna) vemos que son prácticamente 0 para la característica 1 de Materia Blanca.

4.3. Hipótesis nula H_0 : La característica 1 de la clase 3 es gaussiana con un nivel de significancia (α) de 0.001

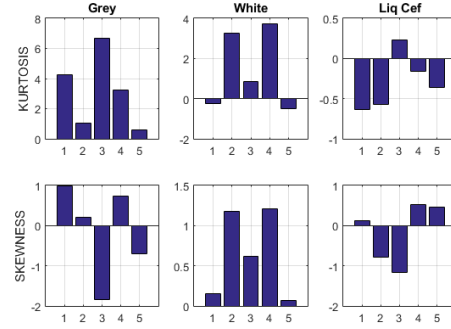
Se mantendrá la hipótesis nula $h = 0$.

El valor de p es $p - value = 6,84e - 14$. Los grados de libertad son 7.

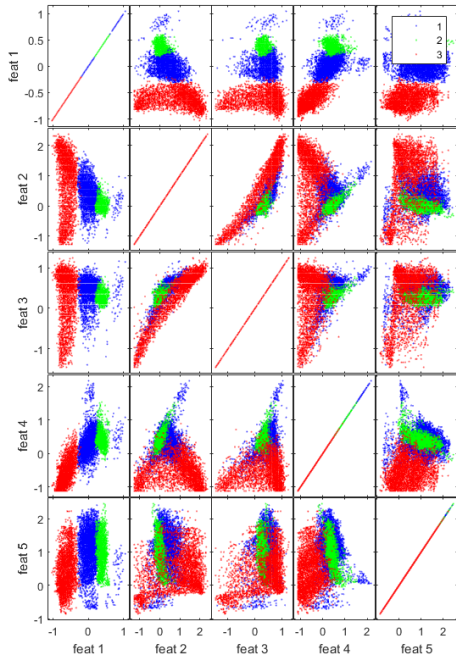
En la figura 2b, 3ra columna, vemos el Skewness y Kurtosis de la característica 1 del líquido cefalorraquídeo. En ella vemos que aunque el Skewness es prácticamente 0, la Kurtosis es demasiado elevada para parecerse a una Gaussiana.



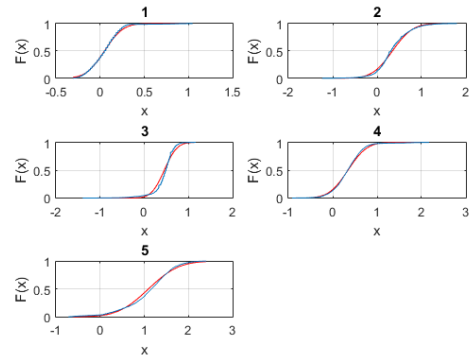
(a) Representación en escala de colores de las características de la base de datos.



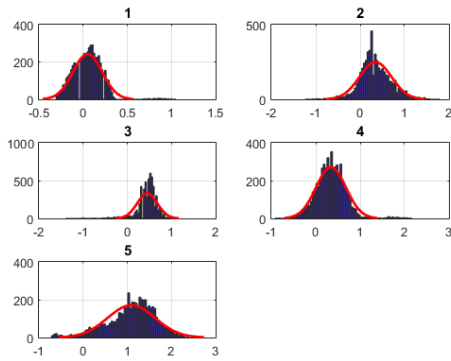
(b) Gráficos de *Skewness* y *Kurtosis*.



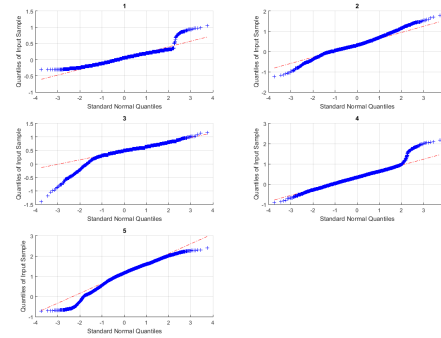
(c) *Scatter plots* de todas las características.



(d) Histogramas acumulados para materia gris

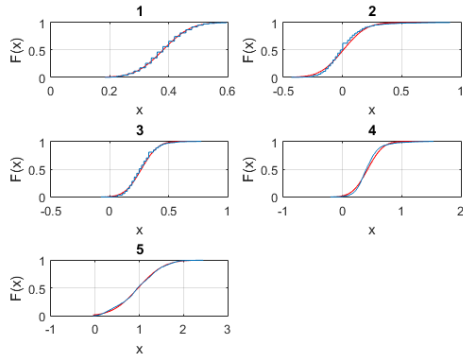


(e) Histogramas para materia gris

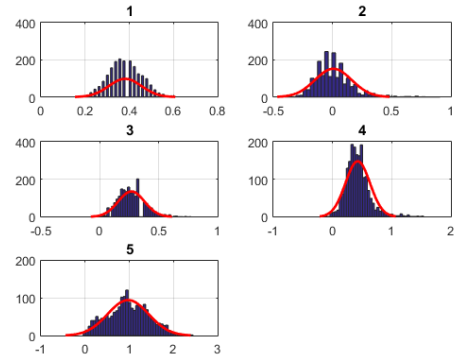


(f) Gráficos *NormPlot* para materia gris

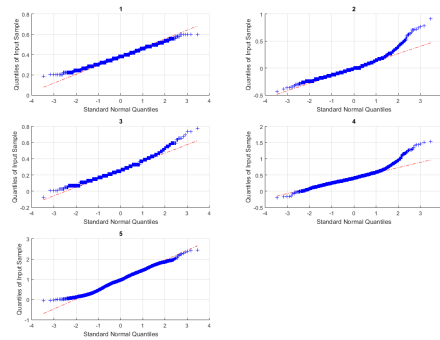
Figura 2: Gráficos de las mediciones sobre las características de materia gris.



(a) Histogramas acumulados para materia blanca

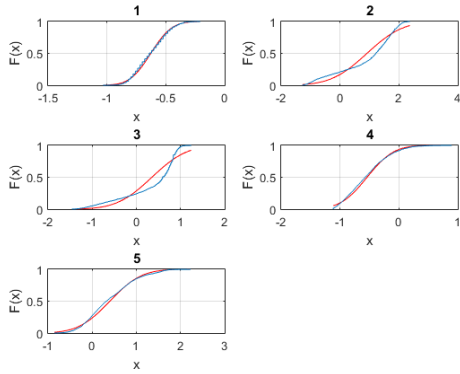


(b) Histogramas para materia blanca

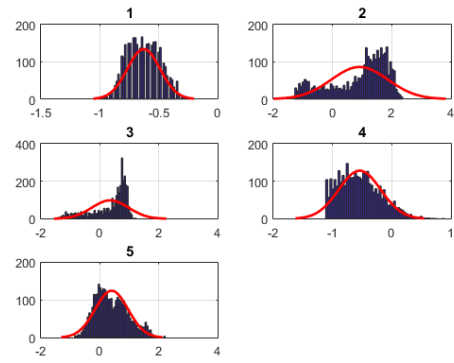


(c) Gráficos *NormPlot* para materia blanca

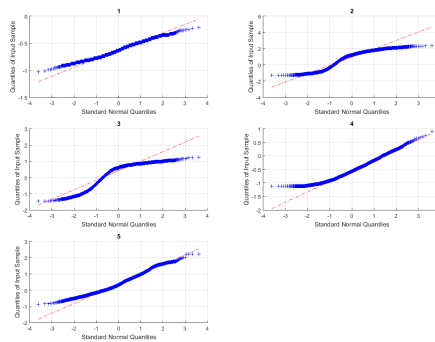
Figura 3: Gráficos de las mediciones sobre las características de materia blanca.



(a) Histogramas acumulados para líquido cefaloraquídeo



(b) Histogramas para líquido cefaloraquídeo



(c) Gráficos *NormPlot* para líquido cefaloraquídeo

Figura 4: Gráficos de las mediciones sobre las características de líquido cefaloraquídeo.