Leyes de potencia

*Tomas Mastantuono(522/23), Juan Milone(264/23)*

[tomastantuono@gmail.com](mailto:tomastantuono@gmail.com) [juanmamilone@gmail.com](mailto:juanmamilone@gmail.com)

*Turno Miércoles 8-14*

# Resumen

Las leyes de escalas o potencias son fundamentales en ámbito de la física debido a que permiten comprender y describir el comportamiento de fenómenos físicos a diferentes escalas. Durante dos etapas, llegaremos a que ciertos fenómenos que en un principio parecen no tener ninguna relación lineal podremos linealizarlos colocando escalas logarítmicas. También veremos como la propia distribución del lenguaje humano, sin importar del idioma, posee una relación que se ajusta de forma muy correcta a la Ley de Zipf y Mandelbrot.

# Introducción

Las leyes de potencia se refieren a patrones matemáticos y estadísticos que describen relaciones proporcionales de potencia entre variables. Estas leyes son ampliamente estudiadas en diferentes campos científicos y se aplican en diversas disciplinas, como física, biología, economía y ciencias sociales.[1]

En este informe de laboratorio, se investigan dos fenómenos relacionados con las leyes de potencia, los cuales van a ser estudiados en dos etapas diferentes, siendo la primera el estudio de la aplicacion de las leyes de escala o potencia sobre la relacion entre hojas de diferentes tamaños y masas, posteriormente la distribucion de uso de palabras en los diferentes lenguajes con textos escritos de dominio publico. Las experiencias recien nombradas van a ser estudiadas y analizadas con las siguientes leyes: *Ley de Zipf* y luego para una segunda etapa, junto con la primera ley, la *Ley de Mandelbrot*. Estas leyes son de gran importancia en diversos campos, ya que describen patrones estadísticos presentes en conjuntos de datos específicos.

La Ley de Zipf, también conocida como la Ley de Rank-Size, es una ley empírica que se encuentra comúnmente en conjuntos de datos de naturaleza lingüística. Esta establece una relación de potencia entre la frecuencia y su rango en un conjunto dado. [2] La Ley de Zipf puede expresarse de la siguiente forma, donde f es la frecuencia, r es el rango y alpha es el parámetro que determina la forma específica de la ley de potencia:

(1)

Por otro lado, la Ley de Mandelbrot, propuesta por el matemático Benoit Mandelbrot, se enfoca en la presencia de patrones fractales en fenómenos naturales y sistemas complejos. Los conjuntos de datos que exhiben una estructura fractal muestran una autosimilaridad a diferentes escalas. La Ley de Mandelbrot revela cómo las propiedades estadísticas de estos conjuntos de datos fractales pueden estar gobernadas por leyes de potencia. [3]

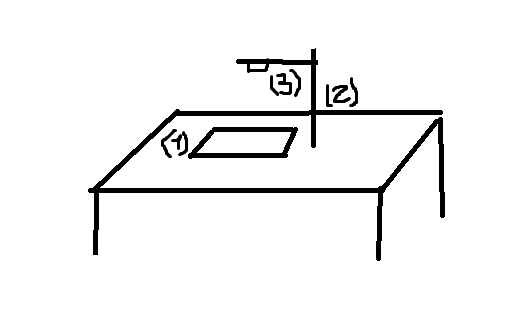
Al relacionar la Ley de Mandelbrot y la geometría fractal con algo tan simple como las hojas, podemos esperar encontrar patrones repetitivos y autosimilares en la estructura de las hojas a diferentes escalas. Además, es posible que encontremos una ley de potencia en las medidas de volumen y perímetro de las hojas, lo que indica la presencia de propiedades fractales. Esta ley puede expresarse de forma similar a la anterior , donde beta es otro parámetro adicional:

(2)

El objetivo de este informe es analizar y comprender la relación entre las leyes de potencia, la Ley de Zipf y la Ley de Mandelbrot, y su aplicación en los experimentos llevados a cabo.

# Desarrollo experimental

El desarrollo de esta experiencia consistió en dos etapas separadas: por un lado el análisis de unas hojas recolectadas del exterior, las cuales poseen distintos tamaños, y de las cuales analizaríamos el cumplimiento de un cierta ley, por otro lado el análisis de textos, historias, y cuentos populares, recolectados de una página web gutenberg.org [4].



*Figura 1: Montaje para la experiencia de las hojas, (1)hoja de papel blanca, (2) palo de laboratorio con soporte para dispositivos móviles, (3)dispositivo movil que utilizariamos como camara de fotos*

El montaje para la experiencia de las hojas consistió en un gancho aferrado a un soporte, para poder sujetar la cámara, como se muestra en la *Figura 1*, y así lograr tomar fotos en un ángulo de 180°, para que se aprecie mejor el volumen que ocupa cada hoja. Bajo la cámara se encontraba una hoja en blanco sobre la que se iban colocando los distintos tipos de hojas.

La experiencia consistió en tomar 4 fotografías para cada uno de los distintos tipos de hojas, que en total fueron 14. Al mismo tiempo, también fuimos midiendo en un balanza (Ohaus Precision Standard, ± 0.05g), el peso de cada una de las hojas (ver en el apéndice, 1), para cada hoja se repite esta medición un total de 4 veces. Es necesario mencionar que debido a la liviandad de alguno de los especímenes recolectados, nos fue necesario agregar a la balanza un peso conocido, de tal forma que genere cierto ruido y podamos obtener algún valor para la masa de la hoja, que de otra forma el display marcaría cero. (este paso no hubiese sido necesario si se hubiera trabajado con un instrumento de medición con más alta precisión).

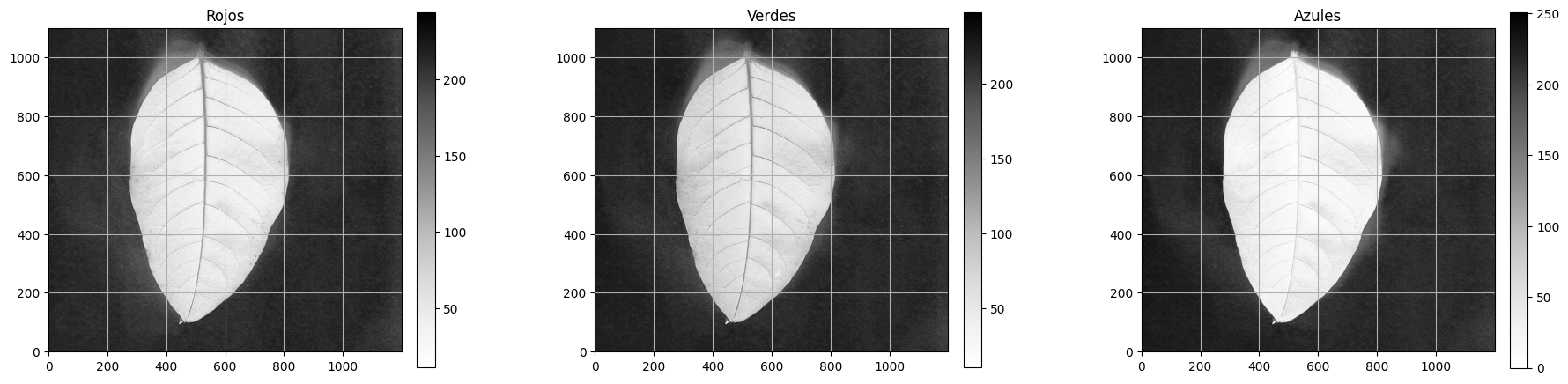
Una vez realizadas todas las mediciones correspondientes, pasamos los datos a python para seguir trabajando con ellos. Con las herramientas de este último programa, logramos realizar un análisis sobre cada una de las fotos. Separamos cada imagenep las componentes de color rojo, azul y verde que la integran.

La segunda parte de la experiencia consistió en el análisis de textos digitales. Para esta etapa no fue necesario el armado de ningún montaje, puesto que todo lo necesario para el cumplimiento de nuestros objetivos y estudio eran escritos los cueales fueron descargados de la página web: *gutenberg.org* [1]. La selección de estas historias, de ahora dominio publico, fueron de forma totalmente arbitraria, junto con los idiomas. La descarga de estos libros fue a través de *plain text UTF*, en un formato *txt*(texto). Una vez obtenidos todos los archivos necesarios, pasamos a un *Google Colab* para hacer el analisis de estos(ver en el apendice, 3).

# Resultados

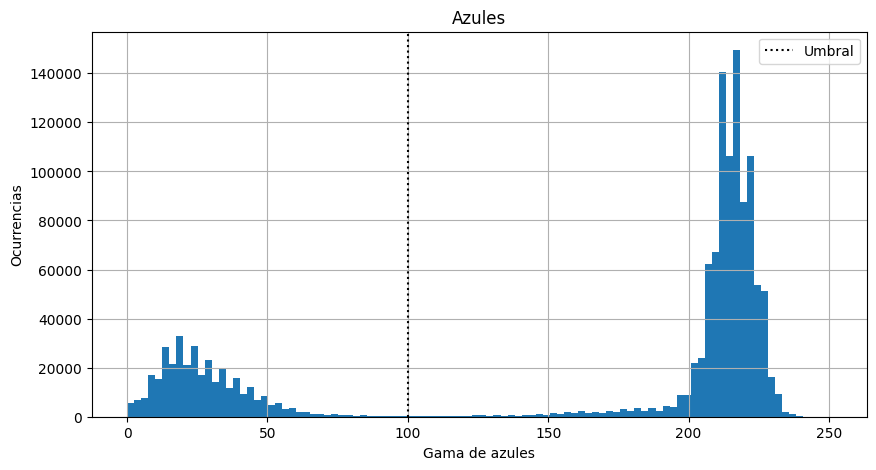
### Áreas y perímetros de las hojas:

Una vez recolectados todos los datos de ambas experiencias comenzamos a realizar un análisis utilizando python.

Para la etapa que consistió en la obtención de imágenes de las hojas de diferentes tamaños y sus diferentes masas, comenzamos con cargarlas dentro del código. Ya con las imágenes subidas decidimos tomar cada foto y separar sus tres dimensiones de colores(red, green, blue), la finalidad de esto era visualizar cuál era el área que tenía cada hoja. Lo mencionado se puede ver en la siguiente figura:

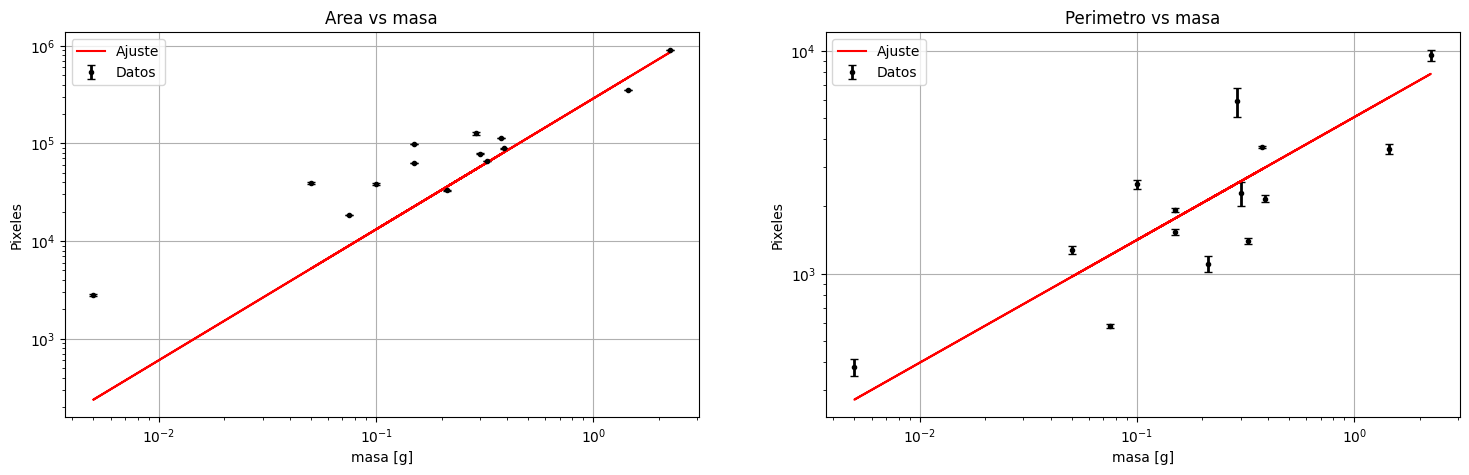
*Figura 2: separación en sus las tres dimensiones RGB*

Ya con las hojas en sus tres colores decidimos utilizar una gama de grises la cual nos permitía visualizar que foto tenía menor cantidad de color de su dimension. La decisión de tomar la dimensión que tenía los colores azules no fue nada arbitraria, ya que este paso era muy importante, porque si tomamos, por ejemplo, la dimensión que tenía todos los verdes de una misma foto, el cálculo del área iba a ser más complicado, debido a que la hoja es mayoritariamente verde.

Para calcular el área de cada hoja, realizamos un plot, como se puede ver a continuación, el cual nos mostraba la cantidad de azules que tenia la foto(apéndice,2):

*Figura 3: Histograma sobre la distribución de los azules, con un umbral el cual marca a partir de cuales pixeles contamos*

El área de la hoja fue calculada a través de contar la cantidad de pixeles que la formaban por la foto. Como se puede ver en la *Figura 3*, colocamos el umbral desde el cual podemos filtrar aquellos píxeles que no formaban parte del área ni el perímetro de la hoja, sino de los alrededores de la imagen.

Una vez obtenidas todas las áreas y perímetros de las hojas y con sus respectivas masas. Decidimos graficar nuestros datos con las áreas y perímetros en el eje de las ordenadas y las masas en el eje de las abscisas. Al momento de realizar el plot, decidimos colocar escalas logarítmicas ya que los datos se encontraban con valores considerablemente altos. Realizamos un ajuste con la *ecuación 1*, y al mismo tiempo una ponderación del mismo sobre lo que habíamos obtenido.

*Figura 4: Plots sobre areas vs masas y perímetros vs masas de la hoja, con sus respectivos ajustes*

Los datos obtenidos a través del ajuste, tanto de las áreas como el de los perímetros son los siguientes:

* Área: ,
* Perímetro:

### 

### 

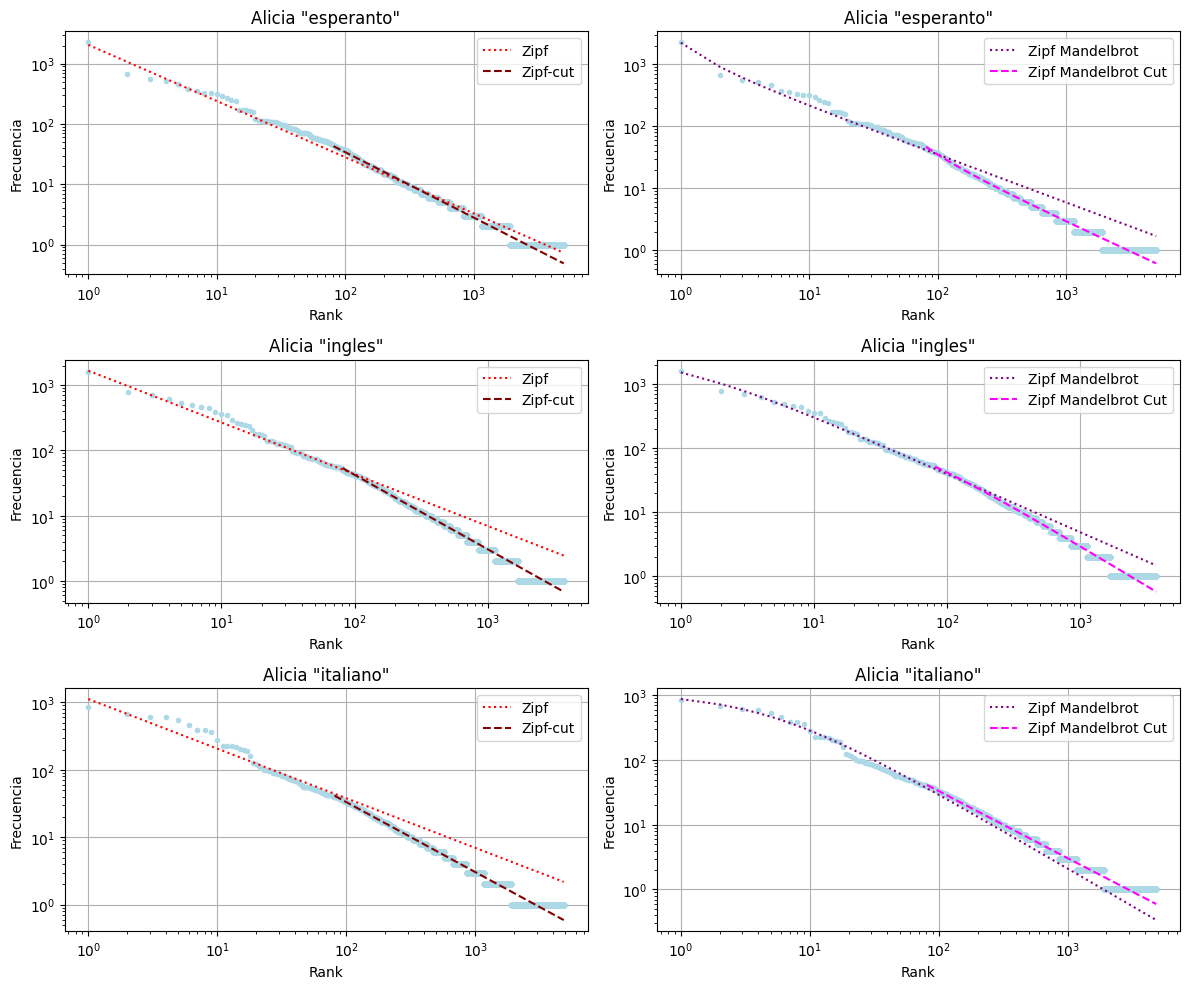
### Distribución de palabras en los lenguajes:

Los libros descargados fueron *Alicia en el país de las maravillas* y *Don Quijote*, siendo en el caso de Alicia descargado en 3 diferentes idiomas(esperanto, inglés e italiano), mientras que Don Quijote fue en 4 diferentes idiomas(catalan, español, finnish y hungaro).

Una vez descargados procedimos a través de código de python, el cual nos filtraba todo el ruido generado por el proyecto gutenberg del cual descargamos los textos, a contar la frecuencia de todas las palabras y categorizar por rangos, con un criterio de que a menor rango mayor frecuencia de la palabra durante el texto.

Ya con las palabras categorizadas y ordenadas, decidimos realizar un plot en el cual colocamos en el eje de las ordenadas a la frecuencia de cada palabra y en el eje de abscisas al rango de aquellas, esta elección fue totalmente arbitraria. Con los plots realizados decidimos realizar diferentes ajustes tanto con la *ecuación 1* como con la *ecuación 2*.

Este proceso fue realizado con los diferentes idiomas de ambos libros mencionados anteriormente:



*Figura 5: en esta figura se puede ver los 3 diferentes idiomas, y por columnas los diferentes ajustes hechos con las leyes de Zipf y la ley de Zipf y Mandelbrot*

Una vez realizados los ajustes vimos que estos no llegaban a cumplir del todo o a la perfección las leyes de escala mencionadas anteriormente, por lo que cuando visualizamos las categorías de palabras vimos que aquellas palabras que tenían mayor frecuencia a ser nombradas era las cuales no tenían un gran significado sino que eran conectores de palabras, tales como ¨de¨, ¨y¨, ¨con¨, etc. Al ver esto decidimos recortar aquellas palabras las cuales dejaban de ser conectores para pasar a ser palabras con un significado más profundo o real, que básicamente no cumplieran con la definición de conectar oraciones, tal como se puede ver en la *Figura 5* con el ajuste de *zip cut*.

Utilizados las ecuaciones ya nombradas obtuvimos del ajuste los siguientes datos:

* mientras que
* mientras que
* mientras que

Finalizado el análisis de para alicia, repetimos todo el proceso pero para Don Quijote con sus 4 idiomas, dándonos los siguientes plots (ver en el apéndice 3)

Luego de visualizar y analizar los gráficos y de también ya haber realizado el ajuste para los 4 idiomas, obtuvimos los siguientes valores:

* mientras que
* mientras que
* mientras que
* mientras que

Vemos que el valor de de alpha en todos los casos nos da muy similar, sin embargo los errores difieren bastante.

# 

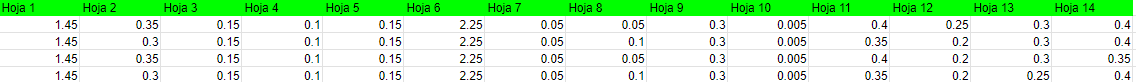
# Conclusiones

Visualizando primeramente en la etapa de la mediciones de áreas y el perímetro de las hojas, pudimos observar que nuestros ajustes no eran para nada buenos en el sentido de la pocas veces el cual este pasa por los datos analizados, sospechamos que esto puede estar ocurriendo por fallos al momento de utilizar las diferentes librerías o de la propia cámara de foto, como también de la decisión de los diferentes umbrales que fuimos colocando como criterio para filtrar a los pixeles que podían o no permanecer a la composición de la hoja. Además no es solo lo visual lo que nos hace darnos cuenta de que el ajuste no es los suficientemente correcto, sino que el chi cuadrado nos da un valor extremadamente grande a comparación de los parámetros libres que tenemos, junto con que si calculamos el p value de estos datos, nos devuelve que este es totalmente cero. Aunque todos estos datos nos mostraran que la ley no se cumple, no podemos arriesgarnos a afirmar esto, ya que es muy posible que se haya producido un error al momento del análisis de datos. También es importante aclarar que este mal ajuste está dado a que los errores en el eje de las ordenadas es muy pequeño pero los puntos se encuentran distancias muy lejanas, siendo muy pocos aquellos que si se mantienen sobre lo que en bases logarítmicas pareciera una función del tipo lineal.

Concluyendo sobre la etapa en la cual analizamos la distribución de palabras en distintos textos en sus variaciones de idiomas, se puede observar fácilmente de los gráficos que todos parecieran tener la misma figura, sin importar del texto ni del idioma. Al momento de realizar los ajustes con la *ecuación 1* y la *ecuación 2*, nos damos cuenta que ninguna de las leyes llegaba a realizar un ajuste el cual pase por la mayoría de los puntos. Al momento de visualizar en qué palabras se realizaba esta especie de quiebre en el ajuste, nos dimos cuenta que era en el punto en el cual las palabras dejaban de cobrar sentido de conectores y se utilizaban para otros fines más cercanos a la definición de oraciones. Esto nos dio otra recta con una pendiente alpha distinta a la que veníamos teniendo. Sin embargo, al realizar esto con los diferentes textos pudimos ver que el comportamiento, sin importar el texto o el idioma, eran aproximadamente iguales, siendo que los valores devueltos por los ajustes, se contenían entre ellos. Creemos que este quiebre del ajuste entre palabras de conectores y de aquellas que no lo son pueden ser dadas por la construcción de un idioma en general de la humanidad, ya que los conectores son muy importantes al momento del desarrollo de una frase, oración o textos.

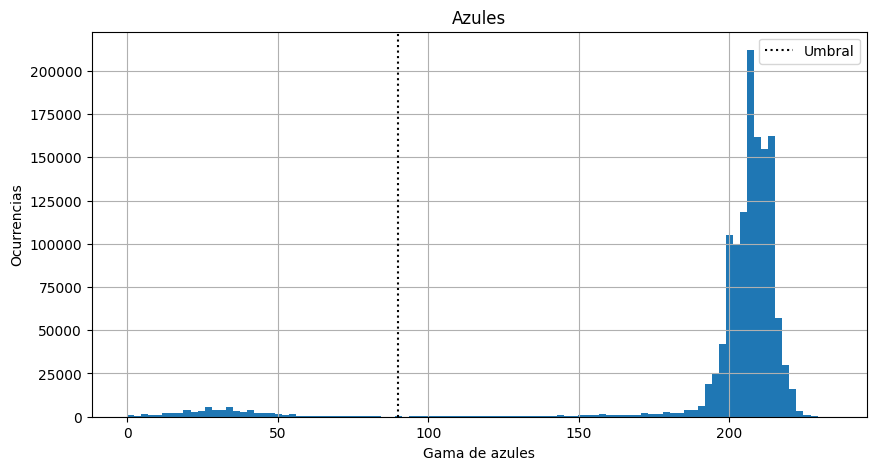
# 

# Apéndice

[1] Masas de las hojas (gr):

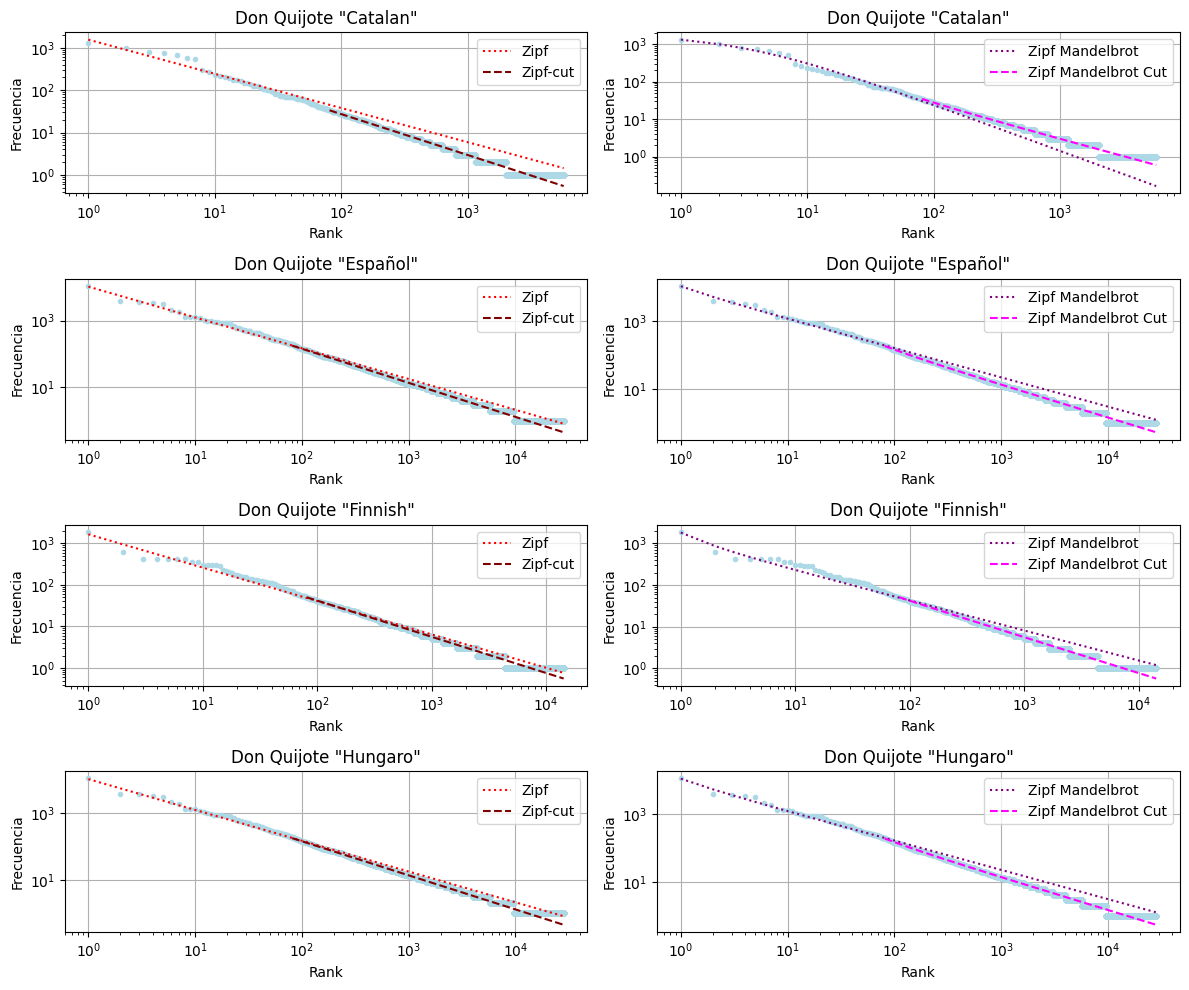
*Figura 7:masas de las diferentes hojas*

[2]Ejemplos sobre distribución de colores



*Figura 8:distribución de la gama de azules*

[3] Figura en la que se puede ver los 3 diferentes idiomas, y por columnas los diferentes ajustes hechos con las leyes de Zipf y la ley de Zipf y Mandelbrot:



# 

# Bibliografía

*[1] Laboratorio 1*, https://media2.utp.edu.co/archivos/Laboratorio%20%200%20(Ley%20de%20escala).pdf. Accessed 27 June 2023.

[2] Rodríguez, Fernando. “¿Qué es la Ley de Zipf?” *KeepCoding*, 15 February 2023, https://keepcoding.io/blog/que-es-y-en-que-consiste-la-ley-de-zipf/. Accessed 27 June 2023.

[3] “Ley de Zipf/Mandelbrot y teoría de la probabilidad aplicadas a la caracterización de reacciones adversas a medicamentos en adultos mayores.” *SciELO Colombia*, http://www.scielo.org.co/scielo.php?script=sci\_arttext&pid=S1794-44492016000200004. Accessed 27 June 2023.

[4] García, Concepción. “Browse By Language: Spanish.” *Project Gutenberg*, https://www.gutenberg.org/browse/languages/es. Accessed 24 June 2023.