

INFORME TAREA 7

Benjamín Briceño
RUT: 20.653.153-3
Github: @Benbriel

1. Problema 1

Esta sección busca estimar el año para el que la anomalía de la temperatura de la Tierra supere los 2 grados celsius. Para esto, se tomaron los datos del GISS, que contienen la diferencia entre la temperatura promedio de 1951 a 1980 y el promedio anual para cada año.

Para esto, se ocupó el algoritmo de minimización de los mínimos cuadrados (*least squares*), que encuentra un mínimo en la suma de cuadrados de las diferencias en el eje y entre los puntos de la función modelada y los datos originales.

Se escogió modelar los datos con una exponencial controlada por 3 parámetros. Esto pues desde hace casi dos siglos, el ser humano ha logrado crecer de forma exponencial, y crear y revolucionar la industria y nuevas tecnologías que generan grandes cantidades de gases de efecto invernadero. Hoy en día, desde hace unos 100 años y, si no cambiamos nada en nuestro comportamiento, en el futuro, el crecimiento exponencial del ser humano en todos los ámbitos generará en el corto plazo un aumento acelerado de la temperatura terrestre. Este comportamiento supera en órdenes de magnitud a los procesos naturales de variación de la temperatura terrestre, por lo que en el corto plazo sigue siendo un simple pero preciso estimador. Con esto, el modelo elegido se comporta tal que

$$T(x) = n + A \cdot e^{r(x-1880)} \quad (1)$$

En este caso, la exponencial se encuentra centrada en el año 1880, considerando que coincide aproximadamente con el inicio de la revolución industrial en la mayoría de países del mundo. Este valor es totalmente arbitrario, pero permite modelar la exponencial con mayor precisión y que el parámetro A no sea tan pequeño. El resto de parámetros le otorgan una mayor libertad al algoritmo de mínimos cuadrados para encontrar un mejor modelo.

Cabe destacar el uso del módulo `lmfit`, que por medio de clases de `Python` permite la creación de parámetros especiales para el modelamiento de datos, otorga una variedad de estimadores y calcula de forma iterativa la incertidumbre de cada parámetro. Además, `lmfit.Minimize.minimize` permite implementar el algoritmo de *least squares* de forma breve.

La Figura 1 muestra la curva exponencial de mejor ajuste a los datos obtenidos. Se obtuvo un valor para los parámetros de $n = -0.314$, $A = 0.0397$ y $r = 0.0249$. Extrapolando esta curva y realizando una búsqueda de raíces, se estimó que la variación de temperatura alcanzará los 2°C aproximadamente en el año 2043. Además, realizando un muestreo de los datos tipo *bootstrap*, se estimó un intervalo de confianza al 68% de [2041.3, 2045.5].

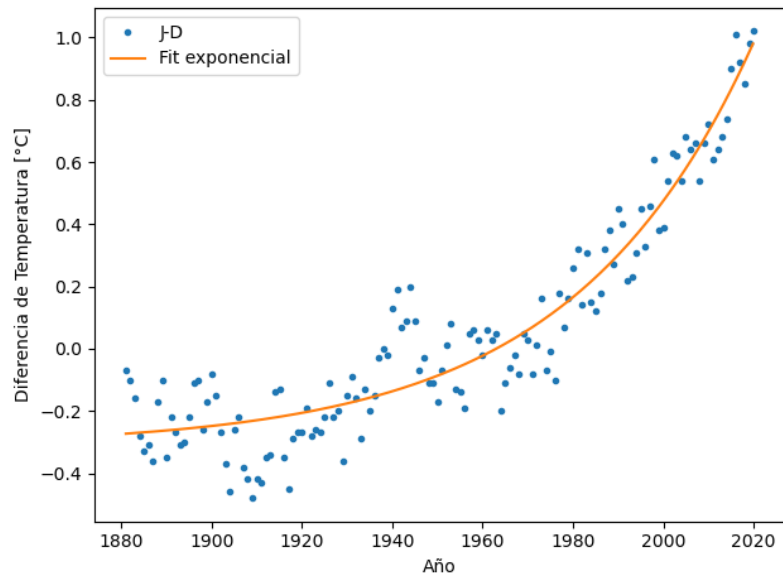


Figura 1: Datos del promedio anual de variación de temperatura J-D (*January - December*) y la curva de mejor ajuste encontrada.

2. Problema 2

2.1. Introducción

Esta parte de la tarea tiene como objetivo el estudio y modelamiento de un segmento del espectro de una fuente de radiación continua con líneas de emisión. Para esto, se planteó un modelo formado por un continuo sumado a dos líneas de emisión equivalentes a curvas gaussianas de anchos iguales. En este caso, el modelo planteado es

$$S(x) = m \cdot x + n + A \cdot \mathcal{N}(\mu_1, \sigma^2, x) + B \cdot \mathcal{N}(\mu_2, \sigma^2, x) \quad (2)$$

Donde \mathcal{N} es una gaussiana de media μ y desviación estándar σ tal que $\mathcal{N}(\mu, \sigma^2, x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Se utilizó el método de mínimos cuadrados para encontrar el mejor ajuste de los parámetros, y se determinaron diversas propiedades estadísticas del modelo planteado. Finalmente, se discutió acerca de la precisión del modelo obtenido.

2.2. Desarrollo

En primera instancia se utilizó el módulo `lmfit` de `Python` para representar cada uno de los 7 parámetros (2 del continuo, y 5 para las gaussianas) como objetos de la clase `lmfit.Parameters`. Este módulo permite la utilización de éstos para el modelamiento de datos, posee una gran variedad de estimadores y calcula de forma iterativa la incertidumbre de cada parámetro.

Luego, se intentó estimar la incertidumbre del continuo, al ser desconocida. Para esto, se tomó una sección de los datos donde las líneas de emisión no afectaban, se modeló ese continuo con una recta, y se obtuvo su desviación estándar asumiendo que cada punto x_i difiere del modelo como una gaussianas de la misma desviación centradas en $S(x_i)$. Además, se midió la precisión de esa desviación comparándola con la obtenida utilizando una fracción de los datos. Este valor, denominado `yerr`, se asumió como la incertidumbre de los datos para todo el espectro, y se utilizó para el método de los mínimos cuadrados. En este caso se obtuvo `yerr = 2.6912e-19`.

Finalmente, se utilizó la función `lmfit.Minimize.minimize` para obtener el valor de los parámetros que minimizara los cuadrados de las distancias en el eje y entre los puntos y el modelo, a partir del array de datos (`ydata - modelo(params, xdata)`) / `yerr`, que varía los `params` hasta lograr los óptimos.

La Figura 2 muestra el mejor ajuste a los parámetros luego de 33 iteraciones dado el método de mínimos cuadrados. Además, se obtuvo a partir de la función `lmfit.report_fit` una serie de valores y estimadores del modelo. De la matriz de covarianza se pudo calcular la incertidumbre de cada parámetro como la raíz de la covarianza de cada parámetro consigo mismo $\sqrt{\text{cov}(p_k, p_k)} = \sqrt{\text{var}(p_k)} = \text{std}(p_k)$.

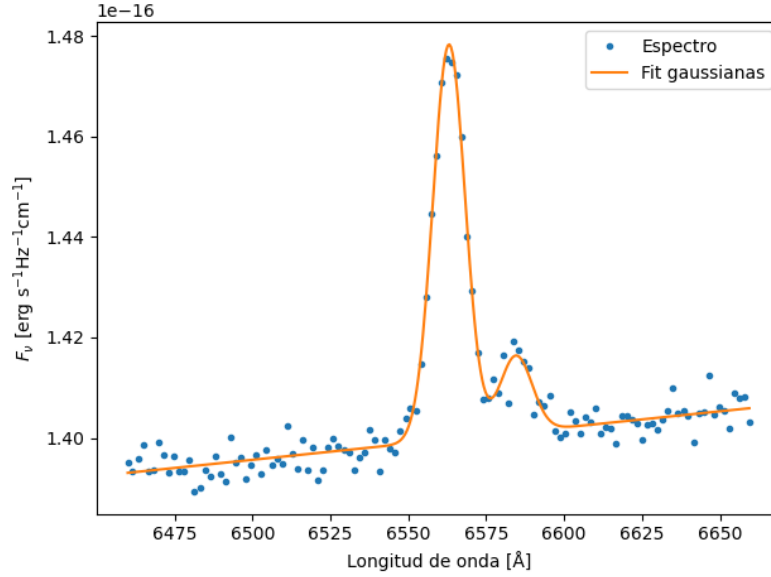


Figura 2: Espectro observado y la curva de mejor ajuste.

Parámetro	Valor de ajuste \pm incertidumbre	Error porcentual (%)
m	$6.4508e-21 \pm 4.3854e-22$	6.80
n	$9.7632e-17 \pm 2.8753e-18$	2.95
A	$1.0211e-16 \pm 1.9966e-18$	1.96
μ_1	6563.1 ± 0.11054	0.00
σ	5.1773 ± 0.10756	2.08
B	$1.9881e-17 \pm 1.6423e-18$	8.26
μ_2	6584.5 ± 0.57732	0.01

Cuadro 1: Parámetros de mejor ajuste y sus incertidumbres. En la tercera columna, el error porcentual muestra incertidumbre / valor (%).

La tabla 1 muestra, además de los valores de mejor ajuste, el error porcentual asociado a cada variable. Éste es significativamente bajo en los valores μ_1 y μ_2 , lo que indica que el algoritmo encontró con precisión los peaks de las gaussianas. Para el resto de parámetros, el error asociado es mayor al 5% para m y B. Esto puede deberse al ruido del continuo, que limita la precisión del cálculo de la pendiente del mismo y la amplitud de la segunda línea de emisión, la cual tiene un tamaño reducido.

Estadística	Valor
χ^2	122.87
χ_{red}^2	1.0684

Cuadro 2: Valores de chi cuadrado y su parte reducida.

La tabla 2 muestra un chi cuadrado reducido de 1.0684, lo que es cercano a la unidad con

diferencia del 6.84%. Esto también puede obtenerse dividiendo χ^2 por la cantidad de grados de libertad, que es el total de datos menos el número de parámetros del modelo ($122 - 7 = 115$).

2.3. Discusión y Conclusiones

El valor de mejor ajuste de los parámetros varía muy poco para los valores de μ_1 y μ_2 , por lo que es altamente probable que el modelo fiteado haya encontrado con precisión dónde se encuentran las líneas de emisión. Sin embargo, para parámetros como B o m existe una incertidumbre de entre el 6% y 8%, lo que indica que puede haber un error en la amplitud de la segunda línea de emisión predicha por el modelo, además de la pendiente del continuo. Esto es de esperar, puesto que existe un ruido en el continuo.

El valor de χ_{red}^2 obtenido permite concluir que una correlación entre los datos medidos para el espectro y el modelo fiteado concuerda con el error calculado en el continuo con una alta probabilidad. Esto, si bien indica que el modelo fiteado tiene una alta probabilidad de modelar bien los datos medidos, no es más que un valor estadístico, y puede verse afectado por una mala suerte de anomalía estadística.

En espectroscopía, el modelamiento de líneas espectrales de ancho y curvas conocidas es bastante efectivo a la hora de medir los parámetros que las describen. Sin embargo, debido a errores de medición o incertidumbres, existe un ruido asociado a los datos medidos. Es la tarea de valores estadísticos como la desviación estándar o χ^2 medir qué tan probable es que un modelo sea aceptablemente preciso, pero la interpretación de estos valores es totalmente arbitraria.