

Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas

FI3104-01 Métodos Numéricos para la Ciencia e Ingeniería

Tarea N°11: Aplicaciones de estadística Bayesiana a ajuste de modelos: Una línea de absorción

Bruno Scheihing, RUT: 18.954.350-6

11 de diciembre de 2015

Resumen

En la presente tarea se buscó modelar una línea de absorción en un espectro mediante técnicas Bayesianas, a partir de dos modelos: un perfil gaussiano sobre el continuo, y la superposición de dos perfiles gaussianos sobre el continuo. Esto consiste en evaluar la probabilidad de que un set de parámetros $\vec{\theta}$ sean los correctos dado una muestra y un modelo. A partir de esto, se logró estimar satisfactoriamente los parámetros óptimos para cada modelo, y sus intervalos de credibilidad al 68 %. Con las densidades de probabilidad calculadas en el espacio de parámetros para cada modelo, se calculó la densidad de probabilidad asociada a que los datos provengan del modelo, y con ello se obtiene el factor Bayesiano, que resultó ser del orden de 10^{63} al dividir la probabilidad de generar los datos con el modelo 2 por la del modelo 1. Es claro desde aquí que el mejor modelo es la superposición de dos perfiles gaussianos sobre el continuo.

1. Introducción

La estadística Bayesiana puede ser una herramienta muy útil para realizar ajustes de modelos, y también para decidir qué modelo representa mejor una muestra de datos. La idea es que dado un modelo M , un conjunto de datos \vec{d} , y un set de parámetros $\vec{\theta}$, la probabilidad de que esos parámetros representen a los datos vía el modelo es:

$$P(\vec{\theta}|\vec{d}, M) = \frac{P(\vec{d}|\vec{\theta}, M) \cdot P(\vec{\theta}|M)}{P(\vec{d}|M)}$$

El término que interesa para determinar qué parámetros son los mejores es el lado izquierdo de la ecuación. El primer término del numerador en el lado derecho se llama verosimilitud, y es una medida de qué tanto se parecen los datos al modelo evaluado en los parámetros $\vec{\theta}$. El segundo es la probabilidad *a priori*, que se asigna directamente a los parámetros, y refleja cuál es la adivinanza inicial para los parámetros. El denominador, por otra parte, se interpreta como una constante de normalización, dado un conjunto suficientemente grande para los parámetros $\vec{\theta}$, tal que la probabilidad de generar los datos sea virtualmente nula afuera de ese conjunto.

En términos más estrictos, trabajaremos con densidades de probabilidad, pues los parámetros provienen de un espacio continuo y los datos también (técnicamente esto no es conocido, pero el modelo a usar lo supone así).

Usando esto, en la presente tarea se busca modelar una línea de absorción en un espectro de flujo por unidades de frecuencia en términos de la longitud de onda. El continuo del espectro es una constante en $10^{-16} [erg\ s^{-1}\ Hz^{-1}\ cm^{-2}]$, y el centro de la línea de absorción está en 6563 Å.

El primer modelo a evaluar consiste en un perfil gaussiano sumado sobre el continuo, pero con signo negativo, de forma de disminuir el valor del flujo y así generar una línea de absorción. Dado que el centro de la línea es conocido, los parámetros libres para este perfil son su ancho σ_1 y su amplitud A_1 .

El segundo modelo consiste en una superposición de dos perfiles gaussianos sumados sobre el continuo (con signo negativo, para poder representar la línea de absorción). Los parámetros libres para ajustar son las amplitudes A_{21} , A_{22} y los anchos de las curvas σ_{21} , σ_{22} .

Se reportarán los mejores parámetros como la esperanza de cada uno de ellos, junto con sus intervalos de credibilidad al 68 %. Se incluirán gráficos para mostrar estos resultados y también gráficos comparativos entre los mejores ajustes de cada modelo y la muestra de datos.

2. Procedimiento

Como los parámetros provienen desde un espacio continuo, y presumiblemente la muestra también, en términos estrictos lo que estaremos calculando está dado por:

$$f_{\vec{\theta}|\vec{d},M}(\vec{\theta}|\vec{d},M) = \frac{f_{\vec{d}|\vec{\theta},M}(\vec{d}|\vec{\theta},M) \cdot f_{\vec{\theta}|M}(\vec{\theta}|M)}{f_{\vec{d}|M}(\vec{d}|M)}$$

donde $f_{X|Y}(x|y)$ representa la densidad de probabilidad condicional de la variable aleatoria X dado un resultado de la variable aleatoria Y .

En ambos casos, supondremos que los errores de la muestra son uniformes sobre todo el espectro, y además de tipo gaussiano. Este será determinado como la desviación estándar de los datos con respecto al continuo, sobre el subconjunto de la muestra que contenga a los puntos suficientemente alejados de la línea de absorción. Dado este error ε , y dado un modelo M descrito por una función g , con su set de parámetros $\vec{\theta}$, se puede calcular la probabilidad de obtener la muestra a partir de:

$$f_{X_i,Y_i|\vec{\theta},M}(x_i,y_i|\vec{\theta},M) = \frac{1}{\sqrt{2\pi\varepsilon^2}} \exp\left(-\frac{(y_i - g(x_i, \vec{\theta}))^2}{2\varepsilon^2}\right)$$

Asumiendo que la muestra es generada independientemente, tenemos que la densidad de probabilidad para generar todo un set de datos $\vec{d} = (\vec{x}, \vec{y})$, con $\vec{x} = (x_1, x_2, \dots, x_N)$ y $\vec{y} = (y_1, y_2, \dots, y_N)$, está dada por:

$$f_{\vec{d}|\vec{\theta},M}(\vec{d}|\vec{\theta},M) = \frac{1}{(2\pi\varepsilon^2)^{N/2}} \exp\left(-\frac{1}{2\varepsilon^2} \sum_{i=1}^N (y_i - g(x_i, \vec{\theta}))^2\right)$$

Por otra parte, para definir la probabilidad *a priori* de los parámetros θ_k de cada modelo, usamos:

$$f_{\theta_k|M}(\theta_k|M) = \frac{1}{\sqrt{2\pi(\delta_k^o)^2}} \exp\left(-\frac{(\theta_k - \theta_k^o)^2}{2(\delta_k^o)^2}\right)$$

donde θ_k^o es el valor esperado *a priori* del parámetro θ_k , y δ_k^o el ancho de la distribución de probabilidad *a priori* asociada. Así, la densidad de probabilidad *a priori* asociada a un set de parámetros θ (de tamaño m) es:

$$f_{\vec{\theta}|M}(\vec{\theta}|M) = \frac{1}{(2\pi)^{m/2} \prod_{k=1}^m \delta_k^o} \exp\left(-\sum_{k=1}^m \frac{(\theta_k - \theta_k^o)^2}{2(\delta_k^o)^2}\right)$$

Los parámetros *a priori* y el ancho de sus distribuciones asociadas serán determinados experimentalmente, por prueba y error, hasta encontrar un set θ^o suficientemente cercano a los mejores parámetros resultantes del cálculo. Para esto basta hacer pruebas “a mano” y verificar gráficamente si los modelos (con esos parámetros) están o no cerca de la muestra \vec{d} .

Con lo anterior, se determina el producto $f_{\vec{d}|\vec{\theta},M}(\vec{d}|\vec{\theta},M) \cdot f_{\vec{\theta}|M}(\vec{\theta}|M)$ a partir del cálculo directo de estas densidades para toda la región relevante del espacio de parámetros, lo que nos interesa pues esto a su vez es proporcional a $f_{\vec{\theta}|\vec{d},M}(\vec{\theta}|\vec{d},M)$. Decimos que un set de parámetros $\vec{\theta}$ está fuera de la región relevante cuando el producto de las densidades de probabilidad es muy cercano a cero, en términos relativos al máximo. Las regiones relevantes se deciden experimentalmente, por prueba y error.

Una vez que se tienen las densidades $f_{\vec{d}|\vec{\theta},M}(\vec{d}|\vec{\theta},M)$ y $f_{\vec{\theta}|M}(\vec{\theta}|M)$ sobre toda la región relevante del espacio de parámetros, se hace una integración del producto entre densidades sobre dicho conjunto. Esta integral es igual a la densidad $f_{\vec{d}|M}(\vec{d}|M)$, por lo que con ello hemos determinado de forma explícita $f_{\vec{\theta}|\vec{d},M}(\vec{\theta}|\vec{d},M)$, desde la primera expresión presentada en esta sección.

Si bien la primera intención es maximizar la densidad de probabilidad, nos conformaremos con la esperanza asociada a la densidad marginal como el valor óptimo para el parámetro. Las densidades marginales para un parámetro dado se pueden calcular a partir de la densidad completa $f_{\vec{\theta}|\vec{d},M}(\vec{\theta}|\vec{d},M)$, integrando sobre todos los demás parámetros.

El cálculo de estas integrales (tanto para calcular $f_{\vec{d}|M}(\vec{d}|M)$ como para el cálculo de las densidades marginales) se realizará multiplicando el valor de la función a integrar en cada punto del espacio de parámetros por el elemento de volumen asociado (que puede corresponder a cualquier número de dimensiones entre 1 y 4, según lo que se esté calculando y el modelo asociado), y luego sumando sobre todo el espacio. El elemento de volumen será uniforme, pues se definirá una grilla uniforme en el espacio de parámetros, permitiendo evaluar la integral de una forma sencilla (aunque no necesariamente óptima): la suma de todos los valores que toma la función en la grilla (o sobre los ejes correspondientes en el caso de las marginales) por el elemento de volumen.

Para cuantificar los intervalos de credibilidad (los intervalos de confianza en estadística Bayesiana) en torno a este valor, se toma el intervalo asociado a cada parámetro que genera un 68 % del área bajo la densidad marginal de probabilidad, y maximiza el mínimo valor que toma la densidad marginal en esta región. En términos prácticos, se parte desde el “centro” de la densidad marginal de probabilidad y se va integrando hacia los costados, procurando que

los valores extremos del intervalo tengan asociado el mismo valor de la densidad de probabilidad, para así asegurar que el intervalo al 68 % buscado esté en la región más probable del parámetro.

Para decidir cuál de los dos modelos es mejor, se procede a comparar las densidades $f_{\vec{d}|M_1}(\vec{d}|M_1)$, $f_{\vec{d}|M_2}(\vec{d}|M_2)$ en los datos \vec{d} , que es el único punto de sus respectivos dominios en donde son conocidas. Tendremos dos casos:

$$\frac{f_{\vec{d}|M_1}(\vec{d}|M_1)}{f_{\vec{d}|M_2}(\vec{d}|M_2)} > 1 \implies M_1 \text{ es mejor} ; \frac{f_{\vec{d}|M_2}(\vec{d}|M_2)}{f_{\vec{d}|M_1}(\vec{d}|M_1)} > 1 \implies M_2 \text{ es mejor}$$

Esto proviene de aplicar el Teorema de Bayes sobre la probabilidad condicional de los modelos dada la muestra, y suponer que *a priori* los modelos son equiprobables.

Durante la resolución numérica, resulta conveniente re-escalar las unidades del problema, de forma tal que los valores numéricos de las cantidades sean más cercanos a 1 ó 10, donde no deberían existir problemas de precisión. Por ello, antes de empezar se amplifican las unidades de flujo por unidad de frecuencia en un factor de 10^{18} . Desde allí, se usa un parámetro de control para encontrar la región más estable donde hacer los cálculos. En las últimas versiones del código, los valores numéricos del flujo por unidad de frecuencia están amplificados por un factor de $2.45 \cdot 10^{18}$, pues este factor hace que el producto $2\pi\varepsilon^2$ (el valor de ε se explicará después) sea suficientemente cercano a 1, de forma tal que al calcular la verosimilitud de los datos, este término no distorsione excesivamente el resultado producto del exponente que lleva (en el set de datos usados, $N = 122$, por lo que $(2\pi\varepsilon^2)^{N/2}$ probablemente será muy grande o muy pequeño).

En lo que sigue, se describe cada modelo a evaluar. Se denotará por c al nivel del continuo, que corresponde a $10^{-16} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}]$, λ a la longitud de onda en el dominio, y λ_0 al centro de la línea de absorción (6563 Å).

2.1. Modelo 1: Línea gaussiana simple

El primer modelo, consistente en una línea gaussiana sobre el continuo, se escribe como:

$$g(\lambda; \vec{\theta}) = c - \frac{A_1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(\lambda - \lambda_0)^2}{2\sigma_1^2}\right)$$

donde $\vec{\theta} = (A_1, \sigma_1)$ son los parámetros a ajustar. Experimentalmente se determinó que los *priors* (parámetros para definir la probabilidad a priori) a usar son:

$$\vec{\theta}^o = (A_1^o, \sigma_1^o) = (8 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}], 3.6 [\text{Å}])$$

Los “anchos” a usar de las distribuciones a priori para los parámetros, elegidos de la misma forma que los valores óptimos, son:

$$\vec{\delta}^o = (\delta_{A_1}^o, \delta_{\sigma_1}^o) = (4 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}], 3 [\text{Å}])$$

Se define una grilla de 201 x 201 puntos para el espacio de parámetros (A_1, σ_1) . Experimentalmente se determinó que un rectángulo definido por el producto cruz (de conjuntos) de los intervalos:

$$A_1 \in (7 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}], 8.3 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}])$$

$$\sigma_1 \in (3.4 [\text{\AA}], 4.0 [\text{\AA}])$$

contiene al conjunto relevante de parámetros. Notar que esta región es más pequeña que la zona donde las probabilidades *a priori* son relevantes, pero en **Resultados** se muestra que el ruido es pequeño, por lo que el término de verosimilitud tiene una sección relevante para los parámetros mucho más pequeña, permitiendo así que el rectángulo antes definido contenga a toda la región de interés del espacio de parámetros.

2.2. Modelo 2: Línea gaussiana doble

El segundo modelo, consistente en dos líneas gaussianas sobre el continuo, se escribe como:

$$g(\lambda; \vec{\theta}) = c - \frac{A_{21}}{\sqrt{2\pi\sigma_{21}^2}} \exp\left(-\frac{(\lambda - \lambda_0)^2}{2\sigma_{21}^2}\right) - \frac{A_{22}}{\sqrt{2\pi\sigma_{22}^2}} \exp\left(-\frac{(\lambda - \lambda_0)^2}{2\sigma_{22}^2}\right)$$

donde $\vec{\theta} = (A_{21}, \sigma_{21}, A_{22}, \sigma_{22})$ son los parámetros a ajustar. Experimentalmente se determinó que los *priors* (parámetros para definir la probabilidad a priori) a usar son:

$$\begin{aligned} \vec{\theta}^o &= (A_{21}^o, \sigma_{21}^o, A_{22}^o, \sigma_{22}^o) \\ &= (4 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}], 2.5 [\text{\AA}], 5 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}], 8.5 [\text{\AA}]) \end{aligned}$$

Los “anchos” a usar de las distribuciones a priori para los parámetros, elegidos de la misma forma que los valores óptimos, son:

$$\begin{aligned} \vec{\delta}^o &= (\delta_{A_{21}}^o, \delta_{\sigma_{21}}^o, \delta_{A_{22}}^o, \delta_{\sigma_{22}}^o) \\ &= (2 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}], 1 [\text{\AA}], 2.5 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}], 3 [\text{\AA}]) \end{aligned}$$

Se define una grilla de 101 x 101 x 101 x 101 puntos para el espacio de parámetros $(A_{21}, \sigma_{21}, A_{22}, \sigma_{22})$. Experimentalmente se determinó que un rectángulo definido por el producto cruz (de conjuntos) de los intervalos:

$$\begin{aligned} A_{21} &\in (2.7 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}], 5.5 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}]) \\ \sigma_{21} &\in (1.7 [\text{\AA}], 3.2 [\text{\AA}]) \\ A_{22} &\in (3.5 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}], 6.2 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}]) \\ \sigma_{22} &\in (6 [\text{\AA}], 11.5 [\text{\AA}]) \end{aligned}$$

contiene al conjunto relevante de parámetros. Notar que esta región también es más pequeña que la zona donde las probabilidades *a priori* son relevantes, al igual que para el primer modelo. La justificación es análoga: como el ruido es pequeño, la verosimilitud acota significativamente la región relevante del espacio de parámetros, permitiendo tomar intervalos relativamente pequeños con respecto a los anchos de las distribuciones de los *priors*.

3. Resultados

A partir de la desviación estándar de la muestra con respecto al continuo, se determinó que una buena estimación para el error (el ruido) asociado a los datos es:

$$\varepsilon = 1.633 \cdot 10^{-19} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}]$$

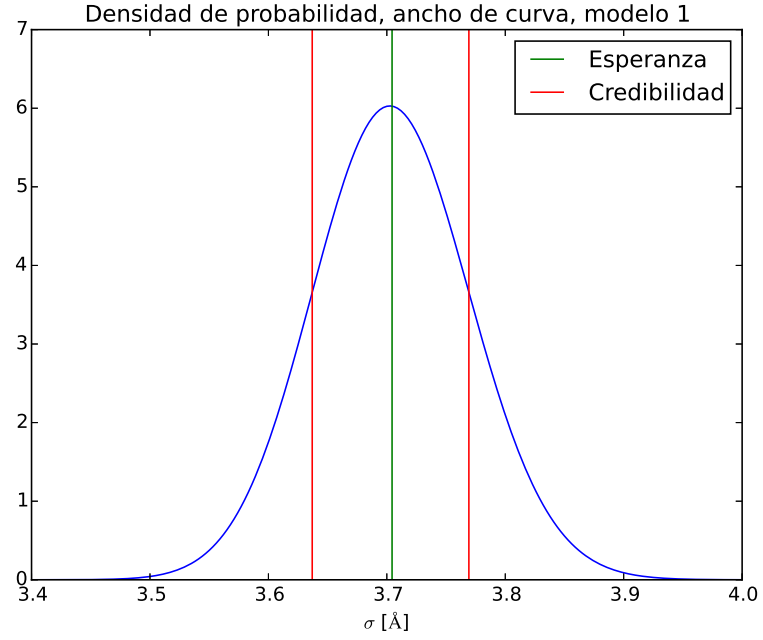
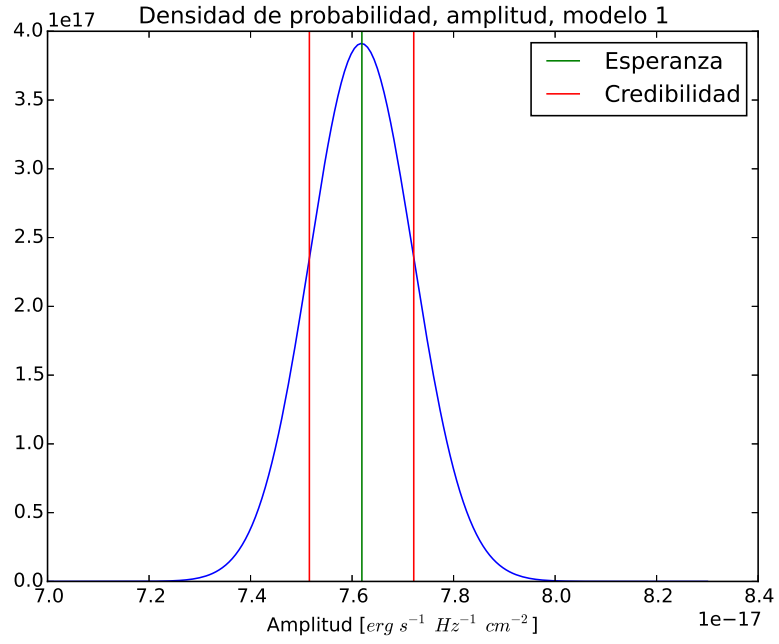


Figura 1: Funciones densidad de probabilidad marginales para parámetros del modelo 1: A_1 y σ

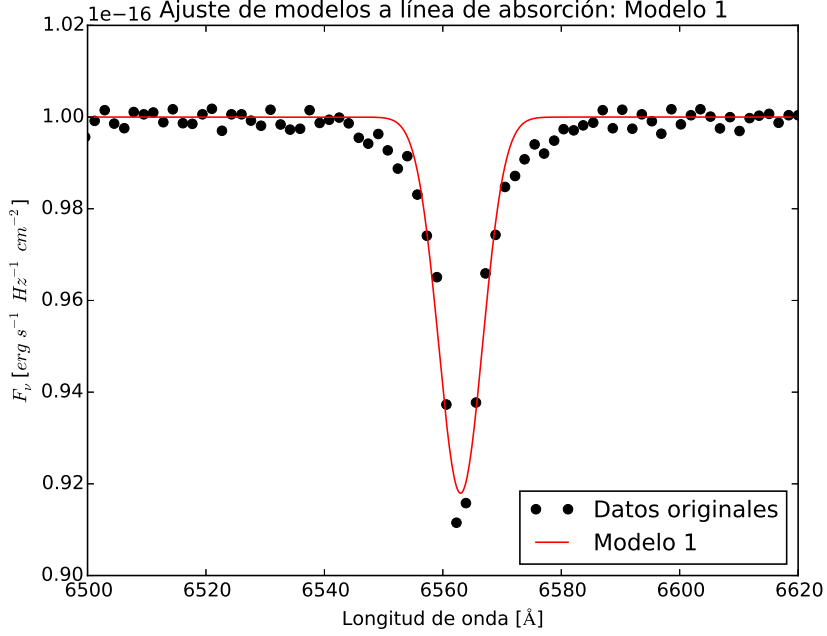


Figura 2: Resultado del ajuste con los mejores parámetros, modelo 1. Los parámetros óptimos son $A_1 = 7.619 \cdot 10^{-17} [erg s^{-1} Hz^{-1} cm^{-2}]$ y $\sigma_1 : 3.704[\text{\AA}]$.

3.1. Modelo 1: Línea gaussiana simple

Al tomar la esperanza de las densidades marginales asociadas a cada parámetro, y hacer el cálculo del intervalo de credibilidad, se obtuvo:

$$A_1 = 7.619 \cdot 10^{-17} [erg s^{-1} Hz^{-1} cm^{-2}]$$

$$\text{Intervalo de credibilidad : } (7.516 \cdot 10^{-17}, 7.726 \cdot 10^{-17}) [erg s^{-1} Hz^{-1} cm^{-2}]$$

$$\sigma_1 : 3.704[\text{\AA}]$$

$$\text{Intervalo de credibilidad : } (3.637, 3.769)[\text{\AA}]$$

En la Figura 1 se muestran gráficamente las densidades marginales de probabilidad obtenidas para cada parámetro. En verde se muestra el valor óptimo, cuantificado por la esperanza, y en líneas rojas se marcan los límites del intervalo de credibilidad. Como se quería, el intervalo de credibilidad toma la zona de mayor densidad de probabilidad.

La Figura 2 muestra un gráfico de los datos originales (dentro de una sección reducida que contiene a la línea de absorción), y del modelo con los mejores parámetros (recién explicitados). Se observa que el modelo falla en reproducir la transición desde el continuo hacia la línea de absorción, pues la curva gaussiana simple se empieza a manifestar más cerca del centro de la línea que la muestra a ajustar. Además, falla ligeramente en reproducir el valor mínimo del espectro, pues la curva del modelo no alcanza dicho valor.

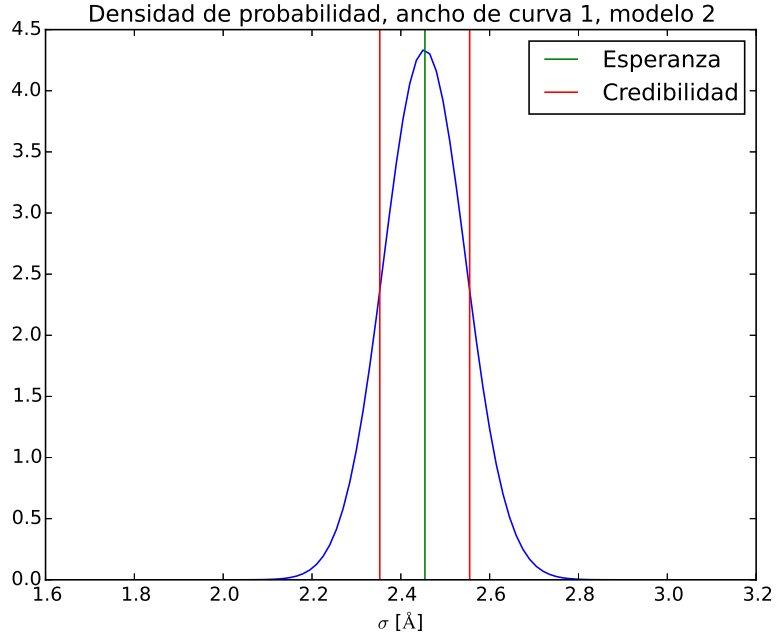
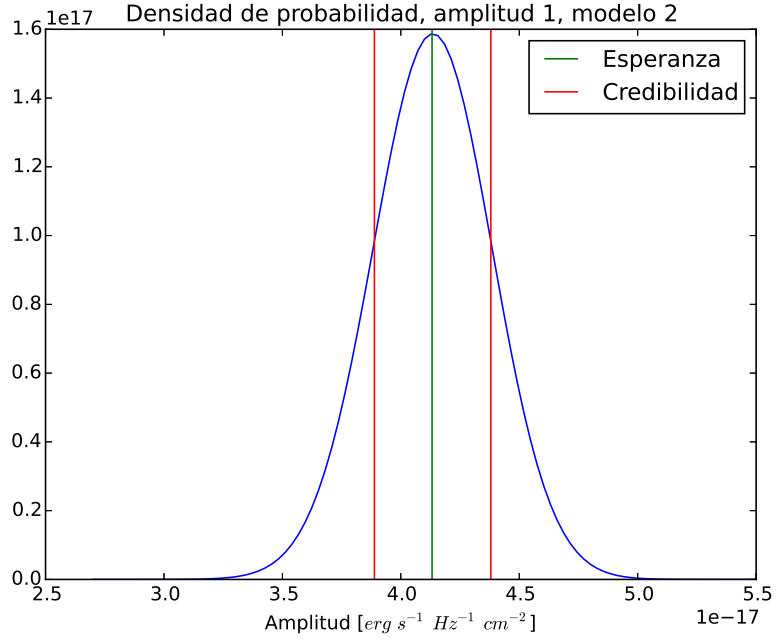


Figura 3: Funciones densidad de probabilidad marginales para parámetros del modelo 2, curva 1: A_{21} y σ_{21}

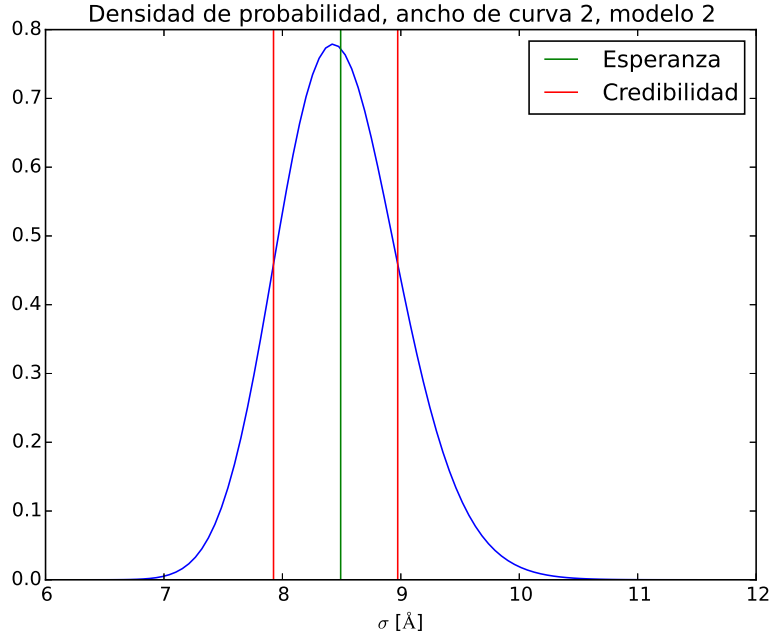
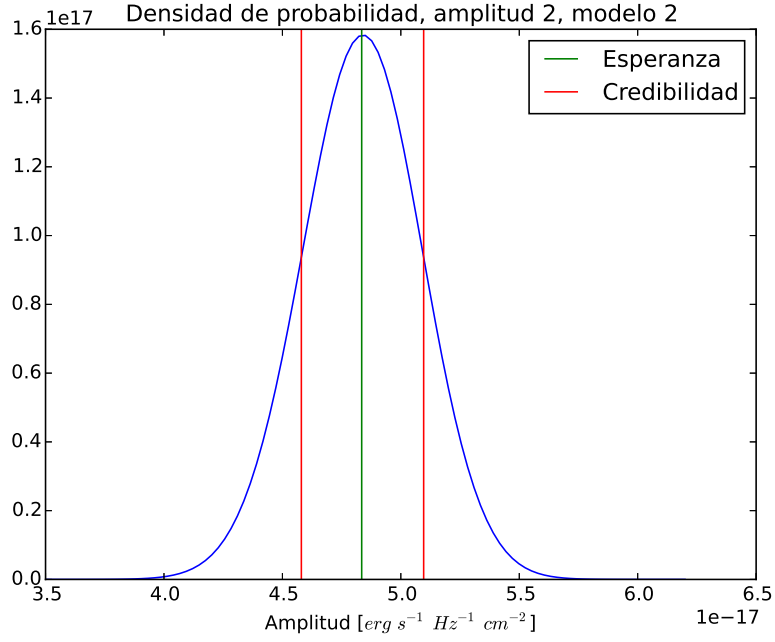


Figura 4: Funciones densidad de probabilidad marginales para parámetros del modelo 2, curva 2: A_{22} y σ_{22}

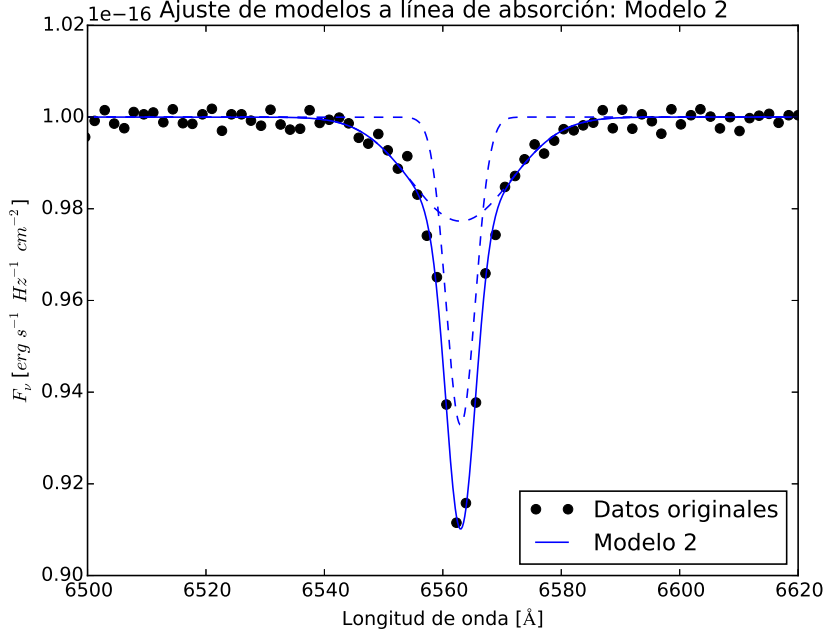


Figura 5: Resultado del ajuste con los mejores parámetros, modelo 2. En línea discontinua se muestra cada gaussiana. Parámetros óptimos en texto adjunto.

3.2. Modelo 2: Línea gaussiana doble

Al tomar la esperanza de las densidades marginales asociadas a cada parámetro, y hacer el cálculo del intervalo de credibilidad, se obtuvo:

$$A_{21} : 4.132 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}]$$

$$\text{Intervalo de credibilidad} : (3.888 \cdot 10^{-17}, 4.380 \cdot 10^{-17}) [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}]$$

$$\sigma_{21} : 2.454 [\text{Å}]$$

$$\text{Intervalo de credibilidad} : (2.353, 2.555) [\text{Å}]$$

$$A_{22} : 4.835 \cdot 10^{-17} [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}]$$

$$\text{Intervalo de credibilidad} : (4.580 \cdot 10^{-17}, 5.096 \cdot 10^{-17}) [\text{erg s}^{-1} \text{Hz}^{-1} \text{cm}^{-2}]$$

$$\sigma_{22} : 8.492 [\text{Å}]$$

$$\text{Intervalo de credibilidad} : (7.925, 8.974) [\text{Å}]$$

En las Figuras 3 y 4 se muestran gráficamente las densidades marginales de probabilidad obtenidas para cada parámetro. En verde se muestra el valor óptimo, cuantificado por la esperanza, y en líneas rojas se marcan los límites del

intervalo de credibilidad. Como se quería (al igual que para el primer modelo), el intervalo de credibilidad toma la zona de mayor densidad de probabilidad.

La Figura 5 muestra un gráfico de los datos originales (dentro de una sección reducida que contiene a la línea de absorción), y del modelo con los mejores parámetros (recién explicitados), con líneas discontinuas para cada gaussiana por separado y en línea continua el modelo completo. Se observa que el modelo representa muy bien los datos, sin perder información de su comportamiento en punto alguno. Esto es una mejora clara respecto del primer modelo.

Para cuantificar cuál modelo es mejor, tenemos que mirar el factor Bayesiano, definido (para este caso) como la división entre las densidades de probabilidad para cada modelo evaluada en los datos. El resultado, luego de calcular las integrales asociadas, es:

$$\frac{f_{\vec{d}|M_1}(\vec{d}|M_1)}{f_{\vec{d}|M_2}(\vec{d}|M_2)} = 2.534 \cdot 10^{-63}$$

Es decir, según esta cuantificación el modelo 2 es mejor por 63 órdenes de magnitud.

4. Conclusiones

Los modelos pudieron ser evaluados exitosamente, obteniendo un set de parámetros óptimos para cada uno utilizando técnicas Bayesianas, y sus intervalos de credibilidad asociados. Visualmente pareciera que los parámetros de cada modelo están efectivamente en el mejor ajuste posible, verificando el buen funcionamiento del método Bayesiano. Además, podemos observar que los intervalos de credibilidad asociados al modelo 2 son más grandes que los asociados al modelo 1, lo que de por sí sugiere cuál modelo tiene más probabilidades de representar a los datos (un intervalo de credibilidad más pequeño sugiere que la verosimilitud disminuye más rápido).

De acuerdo al código implementado, el cálculo de la densidad $f_{\vec{\theta}|\vec{d},M}(\vec{\theta}|\vec{d},M)$ en todos los puntos de la grilla en el espacio de parámetros del modelo 2 toma bastante tiempo, pues hay $101 \times 101 \times 101 \times 101 \approx 10^8$ puntos que evaluar (que quizá no es un número computacionalmente tan grande, pero dada la implementación usada demora bastante). También el espacio en memoria es considerable, pues 10^8 son 100 Megs de información (multiplicado por la cantidad de espacio que usa un bloque de memoria para guardar un número). Resulta claro desde aquí que es posible realizar optimizaciones al código, tanto en tiempo de ejecución como en espacio de memoria utilizado, pues no es necesario guardar todos los valores calculados: interesan solamente su suma y las sumas marginales.

El modelo con dos curvas gaussianas sobre el continuo resultó ser, de acuerdo al factor Bayesiano, escandalosamente mejor que el modelo de una sola curva gaussiana. Esta diferencia se puede atribuir al pequeño error (ruido) asociado a la muestra, pues así si existen puntos en que el modelo está relativamente lejos de la muestra, el factor de verosimilitud se acerca a cero rápidamente. Como el modelo 1 precisamente falla en describir la transición continuo-línea de absorción, perdiendo el comportamiento que exhibe la muestra, es penalizado fuertemente, lo que explicaría la gran mejora que supone el modelo 2 en términos del factor Bayesiano.

También, esta gran diferencia entre modelos sugiere fuertemente que el modelo de dos perfiles gaussianos superpuestos es el mecanismo del cual proviene la muestra. Un test de Kolmogorov-Smirnov permitiría verificar esto con algún grado de certeza.