Métodos numéricos para la ciencia e Ingeniería: Informe Tarea 11

José Guillermo Araya

14 de diciembre de 2015

1. Pregunta 1

1.1. Introducción

El objetivo de la pregunta es aproximar un set de datos que representan una linea de absorción por dos modelos distintos. Primero por una Gaussiana donde los parámetros desconocidos son la amplitud y la varianza, y luego por una suma de Gaussianas (4 parámetros).

Todo esto se realiza usando métodos Bayesianos de estimación.

1.2. Procedimiento

En primer lugar se necesita calcular un error asociado a los datos, este error se estima obteniendo la varianza de los datos en los tramos constantes, es decir:

$$\sigma^2 = \sum (y_{datos} - cte)^2$$

Para el primer fit luego creamos las probabilidades a priori como gaussianas centradas en un valor aproximado para los parámetros, en este caso usamos,

$$\mu_{amplitud} = 1e - 16$$

$$\mu_{varianza} = 100$$

ambas con varianza 100 (prior poco informativo).

Para el prior y la verosimilitud se usa el método visto en clase que utiliza grillas. Para calcular la verosimilitud se calcula como:

$$(2\pi\sigma^2_{datos})^{-N/2} \exp\left(\frac{-\sum(y - modelo(x, param))}{2\sigma^2_{datos}}\right)$$

multiplicando la grilla de las verosimilitudes con la de la prioridad a priori se obtiene la grilla de probabilidad a posteriori. Sumando sobre un eje de esta grilla se obtienen las probabilidades marginales para los parámetros. Finalmente con estas probabilidades se calcula la esperanza del parámetro la cual se usa para realizar el fit.

La probabilidad marginal se normaliza cuidando que la integral total valga 1.

Para el segundo fit, se usa el método de montecarlo para evitar tener que calcular la integral en cuatro dimensiones que demoraría mucho tiempo.

Similar al caso anterior, se define un modelo que consiste en una constante menos 2 gaussianas (de distintas amplitudes y varianzas), por lo que hay que encontrar distribuciones para 4 parámetros.

La estimación inicial de los parámetros es otra vez una normal con promedio el valor esperado del parámetro:

$$\mu_{Amplitud_1} = 1e - 17$$

$$\mu_{Amplitud_2} = 1e - 17$$

$$\mu_{varianza_1} = 8$$

$$\mu_{varianza_2} = 4$$

y nuevamente las varianzas de estas normales son grandes (poca informativas).

Luego se utiliza "find_MAP"de pymc3, para encontrar los valores de los parámetros que maximizan la probabilidad posterior, y estos se usan para graficar el fit,

1.3. Resultados

Realizando el proceso para el primer modelo se obtiene el siguiente fit. donde:

$$Amplitud = 7{,}618 \times 1e - 17 F_\nu [ergs^{-1}Hz^{-1}cm^{-2}]$$

$$Varianza = 3{,}7038035 F_\nu [ergs^{-1}Hz^{-1}cm^{-2}]$$

$$\sigma_{datos} = 0{,}0144135 \times 1e - 17 F_\nu [ergs^{-1}Hz^{-1}cm^{-2}]$$

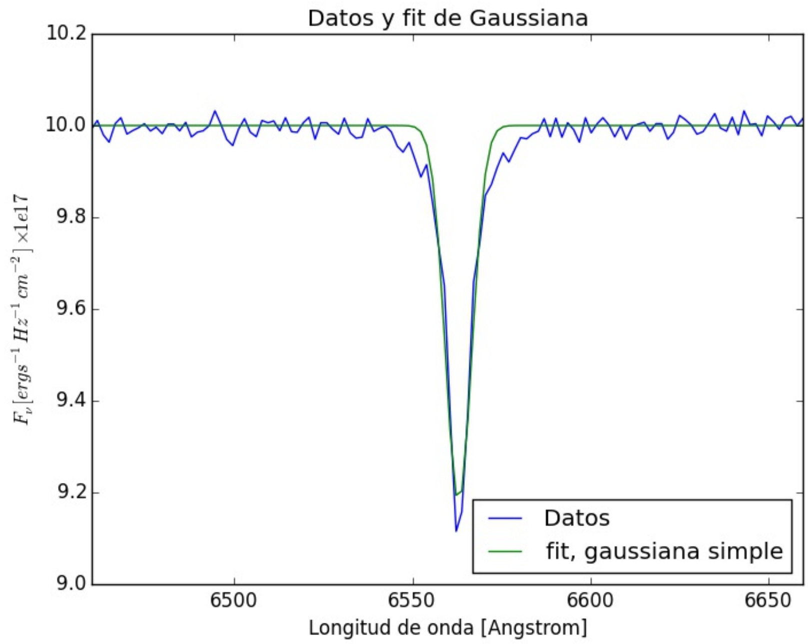

Figura 1: fit de la Gaussiana a los datos

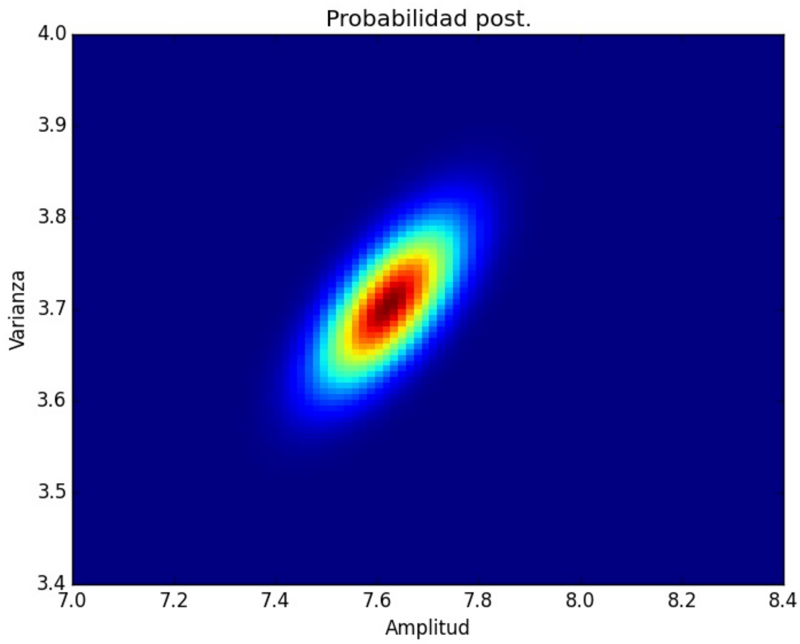

Figura 2: distribución de la probabilidad posterior

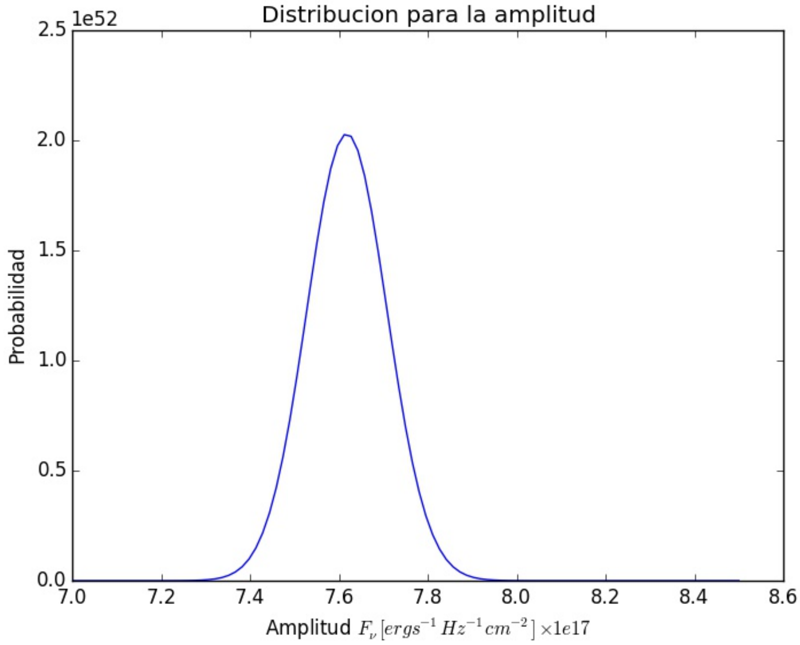

Figura 3: distribución de la probabilidad para la amplitud

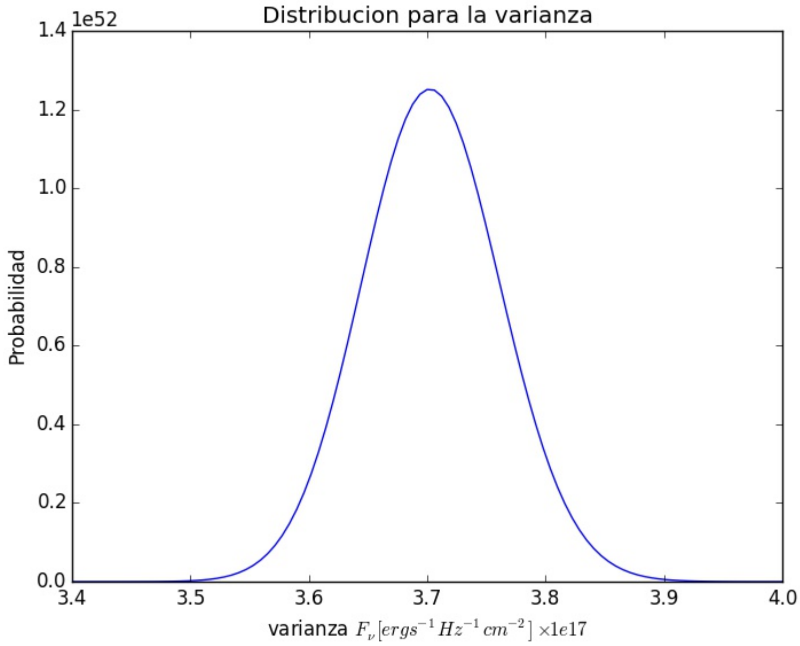

Figura 4: distribución de la probabilidad para la varianza

Para el segundo modelo se obtienen los siguientes valores para los parámetros:

$$Amplitud_1 = 41{,}0314 \times 1e - 18F_\nu[ergs^{-1}Hz^{-1}cm^{-2}]$$

$$Varianza_1 = 2{,}4427F_\nu[ergs^{-1}Hz^{-1}cm^{-2}]$$

$$Amplitud_2 = 48{,}559 \times 1e - 18F_\nu[ergs^{-1}Hz^{-1}cm^{-2}]$$

$$Varianza_2 = 8{,}411F_\nu[ergs^{-1}Hz^{-1}cm^{-2}]$$

y las distribuciones marginales como se observa en la figura 6.

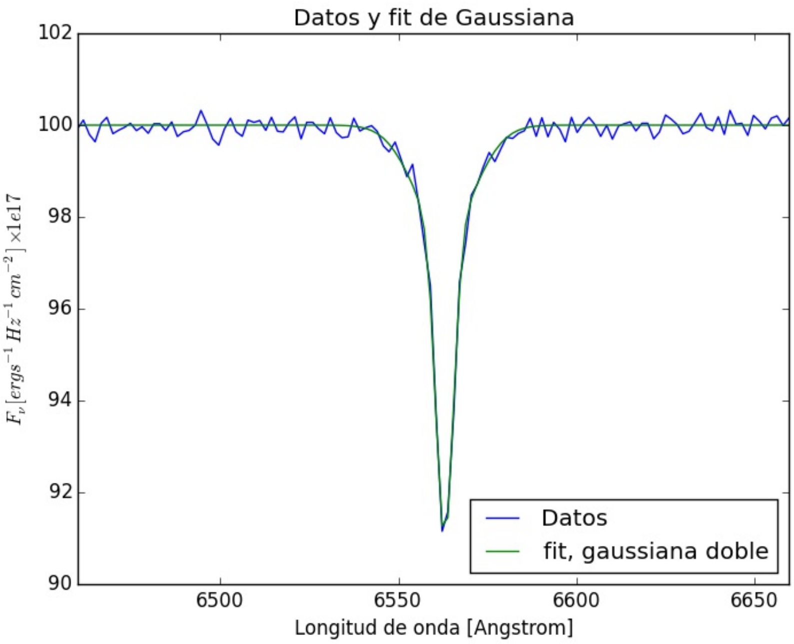

Figura 5: Fit de dos gaussianas a los datos

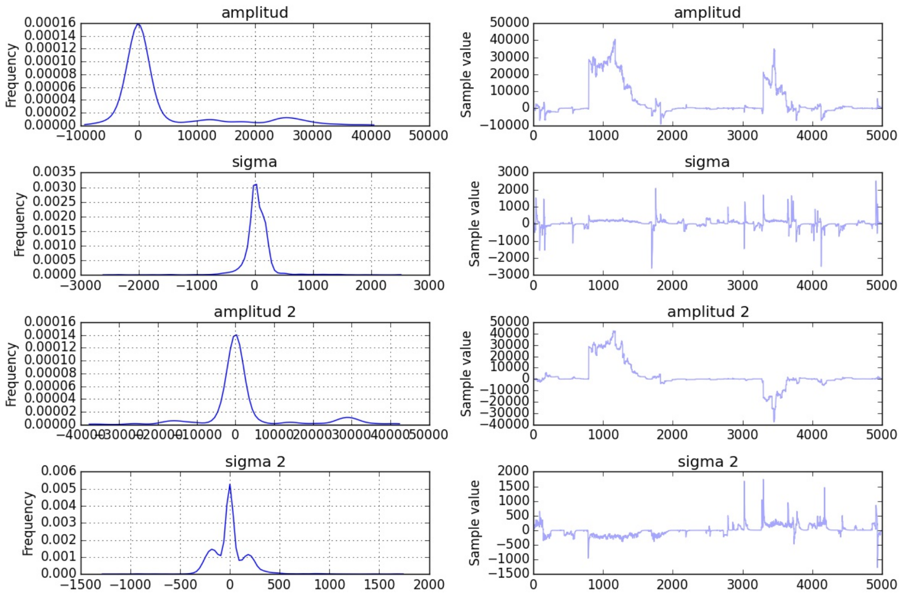

Figura 6: Distribución de probabilidad para los parámetros.

1.4. Conclusiones

Los métodos bayesianos aproximan correctamente los datos, y el segundo fit parece más probable como es esperado (esto se comprobará en la pregunta 2). Las probabilidades a priori usadas tienen información suficiente como para terminar en un modelo apegado a lo esperado, incluso cuando las varianzas de las normales usadas para estas son del orden de 100.

Las gráficos de probabilidades marginales muestran un claro peak, de modo que los parámetros encontrados se aproximan bien por la esperanza de la distribución en el primer modelo.

El segundo modelo representa mejor las .ªlas"de la línea de absorción.

2. Pregunta 2

2.1. Introducción

Esta pregunta consiste en decidir cuál de los dos modelos representa mejor los datos del archivo utilizando métodos bayesianos.

2.2. Procedimiento

Para discriminar entre los modelos se calculará el factor de bayes. Para esto es necesario calcular las evidencias para cada modelo. Para el primer modelo basta con calcular la suma total de la multiplicación del prior con la verosimilitud

Para el segundo modelo, se utilizará importance sampling. tomando puntos random que se asemejen a la distribución y calculando la multiplicación de su prior y su verosimilitud, sumando estos valores para muchos puntos se obtiene el valor aproximado para la evidencia sin tener que realizar una integral en 4-D

Luego se realiza la división de ambo valores y se comparan para determinar que modelo es el más probable.

2.3. Resultados

La evidencia para el primer modelo:

$$E_1 = 2{,}69445739753e + 50$$

Evidencia para el segundo modelo:

$$E_2 = 2{,}39222835958e - 260$$

Factor de bayes:

$$\frac{E_1}{E_2} = 1{,}126338e - 210$$

2.4. Conclusiones

Claramente, el calculo de evidencias tiene un error, ya que se debería obtener el resultado contrario, es decir, el fit de 2 gaussianas debería ser mejor.

Probablemente se deba a un error de escalas, ya que por la magnitud de los flujos con los que se trabajo, requería amplificar los datos por un factor mínimo de 10^{17}. Además para el cálculo de evidencia para el segundo modelo, al usar la gaussiana multivariada, el resultado se indeterminaba por lo que fue necesario escoger desviaciones estándar muy grandes que pueden haber afectado el cálculo.