

# Proyecto Programación básica: Cadenas de Markov Chain Montecarlo (MCMC) con el método de Metropolis.

## Máximo 2 personas por equipo

Prof. Alma González

November 26, 2018

El método de ajuste de parámetros conocido como minimización de  $\chi^2$ , es un caso particular de un proceso de maximización de una distribución de probabilidad. Supongamos que  $y_i^{obs}$  y  $x_i^{obs}$  son el conjunto de puntos colectados de un experimento, y  $\sigma_i$  es el error en cada punto y que aparentemente siguen una recta con ordenada al origen cero. Entonces, para saber cuál es el valor de la pendiente y la ordenada al origen que describe esa recta podríamos proceder a hacer un ajuste de mínimos cuadrados, o mejor aún una minimización de la función  $\chi^2$

$$\chi^2 = \sum_i \frac{(y_i^{obs} - y_i^{mod}(x_i, m, b))^2}{\sigma_i^2}, \quad (1)$$

En general, la pregunta que realmente queremos hacernos es: ¿Cuál es la probabilidad de que un cierto valor de la pendiente, en nuestro modelo de la recta, sea el valor verdadero que describe esos datos. La respuesta la podemos obtener a partir del teorema de Bayes. Si suponemos que el modelo recta debe describir bien los datos, podemos preguntarnos, ¿cuál es la probabilidad de que un valor específico de  $m$  y  $b$ , sean los parámetros que mejor describen los datos observados?. El Teorema de Bayes está dado por:

$$P(\vec{\theta}|\vec{y}) \propto L(\vec{y}|\vec{\theta}) * Pr(\vec{\theta}) \quad (2)$$

Donde  $P$  se denomina distribución de probabilidad a posterior (o Posterior, en inglés),  $L$  se conoce como distribución de probabilidad de verosimilitud (o Likelihood, en inglés) y  $Pr$  se conoce como distribución de probabilidad a priori (o Prior, en inglés). Aquí  $\vec{\theta}$  representa los parámetros que debemos ajustar de nuestro modelo. Encontrar el máximo de la distribución posterior, corresponde a encontrar los parámetros que mayor probabilidad tiene de describir correctamente los datos. Cuando se tienen que el modelo es una recta, los parámetros libres son:  $\vec{\theta} = (m, b)$ , es decir la pendiente y la ordenada al origen.

Lo mas sencillo (pero útil) es suponer una distribución Gaussiana para el Likelihood, que éste relacionada con  $\chi^2$ . Es decir:

$$P(y|m) \propto \exp(-\chi^2/2) \quad (3)$$

En general será mas conveniente trabajar con el logaritmo natural de la probabilidad,

$$\ln(P(y|m)) = -0.5 \sum_i ((y_i^{obs} - y_i^{mod}(x_i, m)) - \sigma_i^2), \quad (4)$$

Por lo que el teorema de Bayes se expresa como:

$$\ln(P(y|m)) \approx \ln(L(\vec{y}|\vec{\theta})) + \ln(Pr(\vec{\theta})) \quad (5)$$

También por simplicidad usaremos una distribución a priori (Prior) plana. Esto es que consideraremos que los parámetros  $m$  y  $b$  tienen igual probabilidad

a priori de describir nuestros datos, siempre y cuando estén en los intervalos  $m_{min} < m < m_{max}$  y  $m_{min} < b < m_{max}$ . Si se cumple que están en dicho intervalo asignaremos  $\ln(Pr(m,b)) = 0$ . (El valor cero es simplemente para indicar que todos tienen igual probabilidad, sería lo mismo asignar cualquier otro valor ya que en nuestro caso no estamos trabajando con las distribuciones propiamente normalizadas.)

El método de Montecarlo maximización de la función posterior más sencillo se conoce como método de Metropolis-Hasting. Consiste en generar los parámetros  $\vec{\theta}$  ( $m$  y  $b$  en este caso) de forma aleatoria y calcular el likelihood y prior para dichos parámetros. Cada vez que se genera un nuevo conjunto de parámetros se debe decidir si se acepta o se rechaza. Dicha decisión depende de si el valor del likelihood es mayor o menor que el valor anterior. A continuación se describen los pasos a seguir.

#### Método de Métropolis

1. Define un conjunto parámetros de inicio:  $m_0, b_0$ . Y calcula el logaritmo del likelihood para estos valores.
2. Genera de forma aleatoria un nuevo conjunto de parámetros:  $m_{new}, b_{new}$ . Estos se generan a partir de una distribución Gausiana centrada en  $m_0$  y  $b_0$ , respectivamente, para lo cual también se debe definir una cierta dispersión  $\sigma_m, \sigma_b$  (se deben hacer varias pruebas de valores de  $\sigma_m$  y  $\sigma_b$ )
3. Calcula el Prior para este conjunto de parámetros. i.e. solamente verifica que  $m_{min} < m < m_{max}$  y  $m_{min} < b < m_{max}$
4. Calcula el valor del logaritmo del likelihood para  $m_{new}, b_{new}$ .
5. Compara el valor del likelihood de los parámetros de inicio con los nuevos. Si los nuevos tienen un valor más positivo, significa que ajustan mejor a los anteriores, con lo cual deben ser aceptados. De aquí se vuelve al punto 1 donde remplazaremos los valores de  $m_0$  y  $b_0$  por lo de  $m_{new}$  y  $p_{new}$ . Si no fue el caso, entonces seguir al siguiente punto.
6. Se debe generar un numero aleatorio entre 0 y 1, y comparar dicho numero con  $np.exp(-\ln(L))$ . Si el numero aleatorio es mayor a  $np.exp(-\ln(L))$  entonces descartaremos los nuevos parámetros  $m_{new}, p_{new}$ . Si el numero aleatorio es menor a  $np.exp(-\ln(L))$  entonces aceptaremos los nuevos parámetros  $m_{new}, b_{new}$ , y volveremos al punto 1 donde remplazaremos los valores de  $m_0$  y  $b_0$  por lo de  $m_{new}$  y  $b_{new}$ .

Los pasos anteriores deben repetirse  $N$  veces, donde  $N$  lo define el usuario. Nota todos los valores  $m_0, b_0$  deben irse escribiendo a un archivo, junto con el valor del likelihood que les corresponde. Habrá casos en los que se repitan los valores debido a que las propuestas no son aceptadas. Dicho archivo será útil para reportar los resultados. Este procedimiento genera lo que llamaremos cadena. Es decir son la secuencia de parámetros que son aceptados por el algoritmo de metrópolis.

Una vez se tienen las cadenas se debe hacer una análisis sobre ellas.

1. Gráficar la secuencia que siguen los parámetros en la cadena.

2. Graficar los parámetros, i.e.  $m$  Vs  $b$ .
3. Graficar un histograma para cada uno de los parámetros.
4. Obtener el valor medio, promedio, y dispersión de los parámetros a partir de dichos histogramas.
5. Graficar el modelo (recta) que corresponde a los valores medios encontrados en el punto anterior, y comparar con los puntos observados.

## 1 Evaluación:

- (5 puntos ) Programa funcionando y evolucionando las partículas. El programa debe hacer uso de funciones, de múltiples archivos para la organización de todo el programa, tener un archivo de compilación (makefile), y usar la mayor cantidad de conceptos vistos en clase (ej. uso dinámico de la memoria). Se obtendrá mejor puntaje cuanto mas organizado, y bien diseñado esté el programa.
- (3 puntos) Presentación de resultados. Incluye: Reporte, Algoritmo, Visualización, Histogramas, etc. Para el reporte se debe investigar sobre el uso de métodos Montecarlo.
- (1punto) Evaluación del equipo
- (1punto) Evaluación individual del profesor
- (1 punto extra) Hacer que el mismo programa pueda usarse para realizar varias cadenas, cada una empezando en un punto  $m_0, b_0$  diferente. El análisis final deberá usar todas las cadenas hechas (es suficiente con 4 cadenas)