## Primer paso para la inferencia Bayesiana: Ajustando una línea recta con Python

Luis Enrique Padilla Albores

**CINVESTAV** 

30/07/2018

$$P(\theta, H|D) = \frac{P(\theta, H)P(D|\theta, H)}{P(D)}$$
(1)

Teorema de Bayes:

$$P(\theta, H|D) = \frac{P(\theta, H)P(D|\theta, H)}{P(D)}$$
(1)

▶  $P(\theta, H|D)$  es el posterior.- Probabilidad de que nuestro modelo (o parámetros) sean correctos según los datos.

$$P(\theta, H|D) = \frac{P(\theta, H)P(D|\theta, H)}{P(D)}$$
(1)

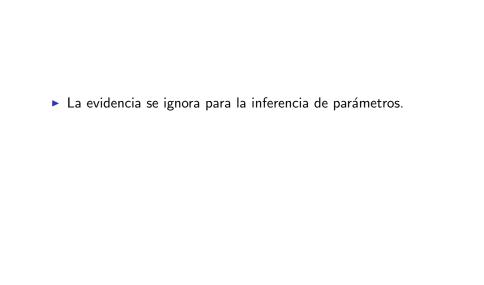
- ▶  $P(\theta, H|D)$  es el posterior.- Probabilidad de que nuestro modelo (o parámetros) sean correctos según los datos.
- $\triangleright$   $P(\theta, H)$  es el prior.- Mide nuestra ignorancia.

$$P(\theta, H|D) = \frac{P(\theta, H)P(D|\theta, H)}{P(D)}$$
(1)

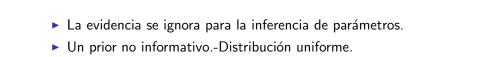
- ▶  $P(\theta, H|D)$  es el posterior.- Probabilidad de que nuestro modelo (o parámetros) sean correctos según los datos.
- $\triangleright$   $P(\theta, H)$  es el prior.- Mide nuestra ignorancia.
- ▶  $P(D|\theta, H) \equiv L(\theta, H)$  es el Likelihood.-Mide la probabilidad de que los datos se ajusten al modelo (parámetros)

$$P(\theta, H|D) = \frac{P(\theta, H)P(D|\theta, H)}{P(D)}$$
(1)

- ▶  $P(\theta, H|D)$  es el posterior.- Probabilidad de que nuestro modelo (o parámetros) sean correctos según los datos.
- $\triangleright$   $P(\theta, H)$  es el prior.- Mide nuestra ignorancia.
- ▶  $P(D|\theta, H) \equiv L(\theta, H)$  es el Likelihood.-Mide la probabilidad de que los datos se ajusten al modelo (parámetros)
- ▶ P(D) es la evidencia



La evidencia se ignora para la inferencia de parámetros.
▶ Un prior no informativoDistribución uniforme.



Un Likelihood usual.-Likelihood Gaussiano.

La evidencia se ignora para la inferencia de parámetros.

▶ Una vez que se obtiene la distribución posterior, se pueden obtener regiones de confidencia de nuestros parámetros.

- Un prior no informativo.-Distribución uniforme.
- Un Likelihood usual.-Likelihood Gaussiano.

Cuando se tienen muchos conjuntos de datos $\rightarrow$  Método de hiperparámetros.

Cuando se tienen muchos conjuntos de datos $\rightarrow$  Método de hiperparámetros.

▶ Si se tienen una colección de datos  $\{D_1, ..., D_N\}$  viniendo de diferentes fuentes  $\{S_1, ..., S_N\}$  (por ejemplo en Cosmología podrían ser supernovas, CMB, etc.), no se sabe a priori si los datos son consistentes entre ellos.

Cuando se tienen muchos conjuntos de datos $\rightarrow$  Método de hiperparámetros.

- ▶ Si se tienen una colección de datos  $\{D_1, ..., D_N\}$  viniendo de diferentes fuentes  $\{S_1, ..., S_N\}$  (por ejemplo en Cosmología podrían ser supernovas, CMB, etc.), no se sabe a priori si los datos son consistentes entre ellos.
- Una forma de obtener información sobre su consistencia es agregando un peso de probabilidad para cada conjunto de parámetros.

Cuando se tienen muchos conjuntos de datos $\rightarrow$  Método de hiperparámetros.

- ▶ Si se tienen una colección de datos  $\{D_1, ..., D_N\}$  viniendo de diferentes fuentes  $\{S_1, ..., S_N\}$  (por ejemplo en Cosmología podrían ser supernovas, CMB, etc.), no se sabe a priori si los datos son consistentes entre ellos.
- Una forma de obtener información sobre su consistencia es agregando un peso de probabilidad para cada conjunto de parámetros.
- ► El método puede ser encontrado en: Lahav O., Bridle S. L., Hobson M. P., Lasenby A. N., & Sodre L., 2000, MN-RAS, 315, 45 o Hobson M. P., Bridle S. L., & Lahav O., 2002, MNRAS, 335, 377 (HBL)

 Se extiende el espacio de parámetros y se utiliza el teorema de Bayes

$$P(\theta, H|D) = \frac{1}{P(D)} \int P(D|\theta, \alpha, H) P(\theta, \alpha, H) d\alpha \qquad (2)$$

 Se extiende el espacio de parámetros y se utiliza el teorema de Bayes

$$P(\theta, H|D) = \frac{1}{P(D)} \int P(D|\theta, \alpha, H) P(\theta, \alpha, H) d\alpha \qquad (2)$$

► El nuevo Likelihood considerando hiperparámetros y partiendo de un Likelihood Gaussiano es

$$P(D; \theta, H_1) = \prod_{k=1}^{N} \frac{2\Gamma(\frac{n_k}{2} + 1)}{\pi^{n_k/2} |V_k|^{1/2}} (\chi_k^2 + 2)^{-(\frac{n_k}{2} + 1)}$$
(3)

Es posible obtener cuando un método de inferenia es mejor que el otro. Considerando la razón

$$K = \frac{P(D|H_1)}{P(D|H_0)} \tag{4}$$

Es posible obtener cuando un método de inferenia es mejor que el otro. Considerando la razón

$$K = \frac{P(D|H_1)}{P(D|H_0)} \tag{4}$$

Se obtiene de forma explícita

$$K = \prod_{k=1}^{N} \frac{2^{n_k/2 + 1} \Gamma(n_k/2 + 1)}{\chi_k^2 + 2} e^{-\chi_k^2/2}.$$
 (5)

Es posible obtener cuando un método de inferenia es mejor que el otro. Considerando la razón

$$K = \frac{P(D|H_1)}{P(D|H_0)} \tag{4}$$

Se obtiene de forma explícita

$$K = \prod_{k=1}^{N} \frac{2^{n_k/2+1} \Gamma(n_k/2+1)}{\chi_k^2 + 2} e^{-\chi_k^2/2}.$$
 (5)

K value	Strenght of evidence
< 1	Negative (supporting $H_0$ )
1 to 3	Weak
3 to 10	Substantial
10 to 30	Strong
30 to 100	Very Strong
> 100	Decisive

Table 1: Criteria for the ratio between two Bayesian evidence, taken from [?].

#### Herramientas numéricas

- ▶ Se utiliza un MCMC para el cálculo del posterior.
- Se escoge un criterio de aceptación (o no) del paso.-Metropolis Hasting
- Es necesario un criterio de convergencia.- Gelman-Rubin
- Dibujar regiones de confidencia para los parámetros que mejor se ajustan a los datos.
- Usaremos PyMC3 para resolver nuestro primer ejemplo
- PyMC3 (https://docs.pymc.io/) es un paquete de Python que nos permite hacer inferencias bayesianas a partir de métodos avanzados de MCMC.