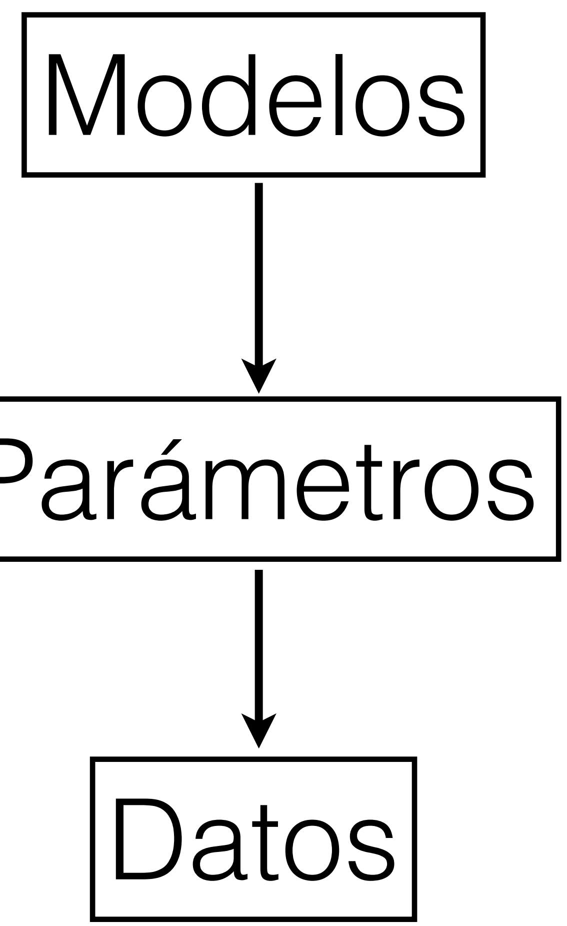
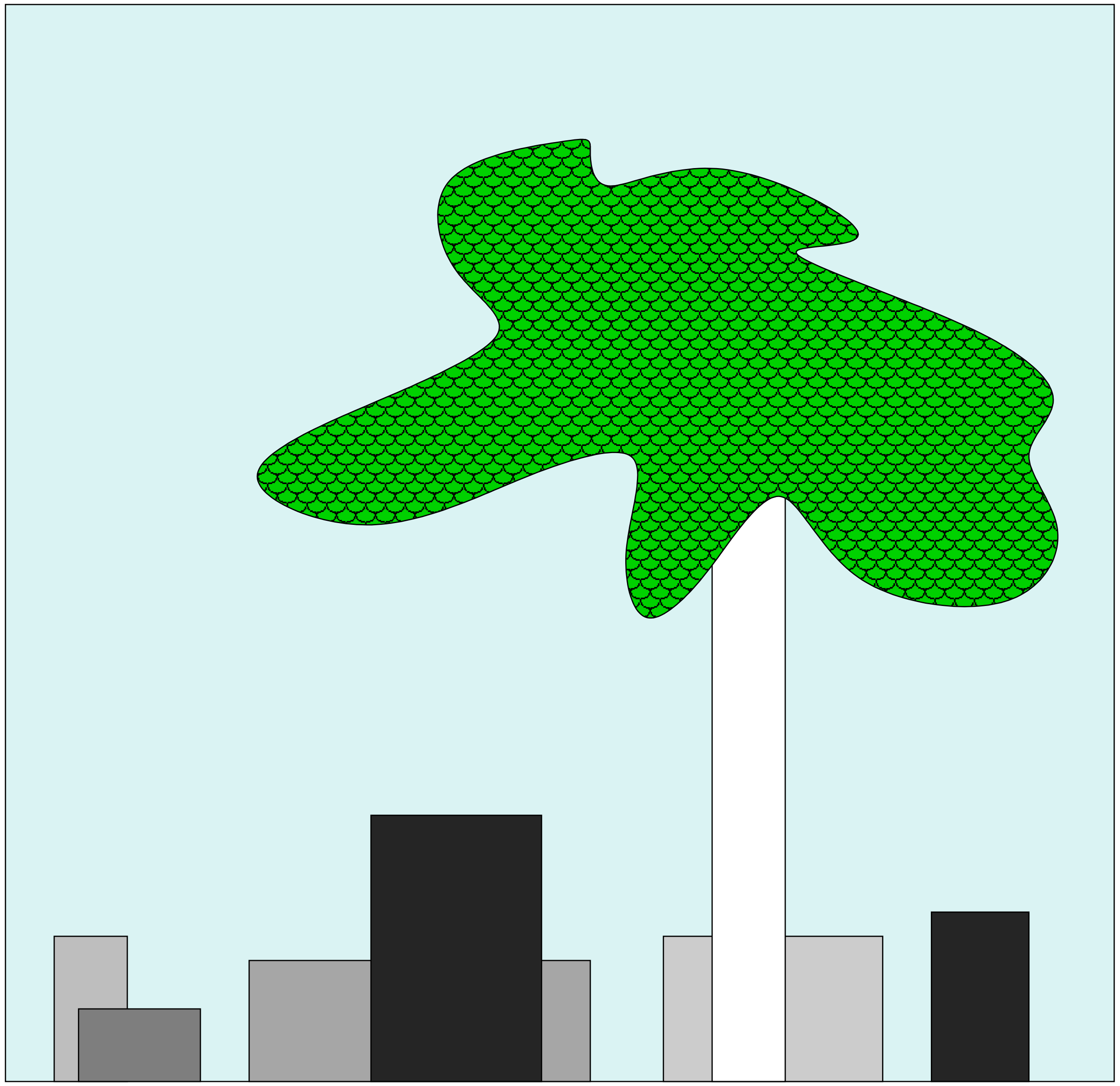
Fundamentos de Inferencia Bayesiana

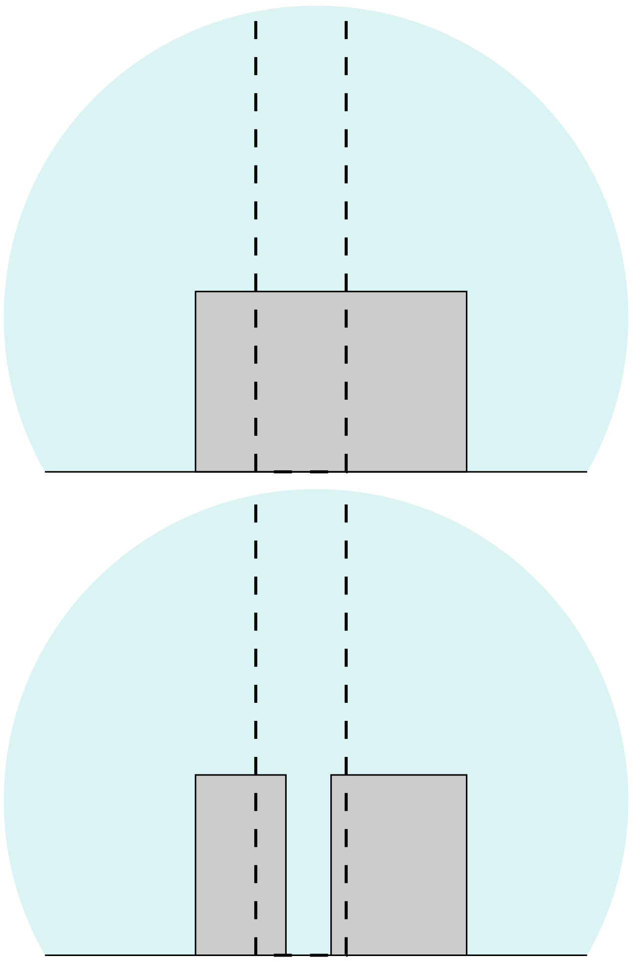
Comparación de Modelos

Comparación de Modelos y la “Navaja de Occam”



En principio, un problema jerárquico





¿Cuántas cajas hay atrás del árbol?

Occam: prefiramos la explicación más simple

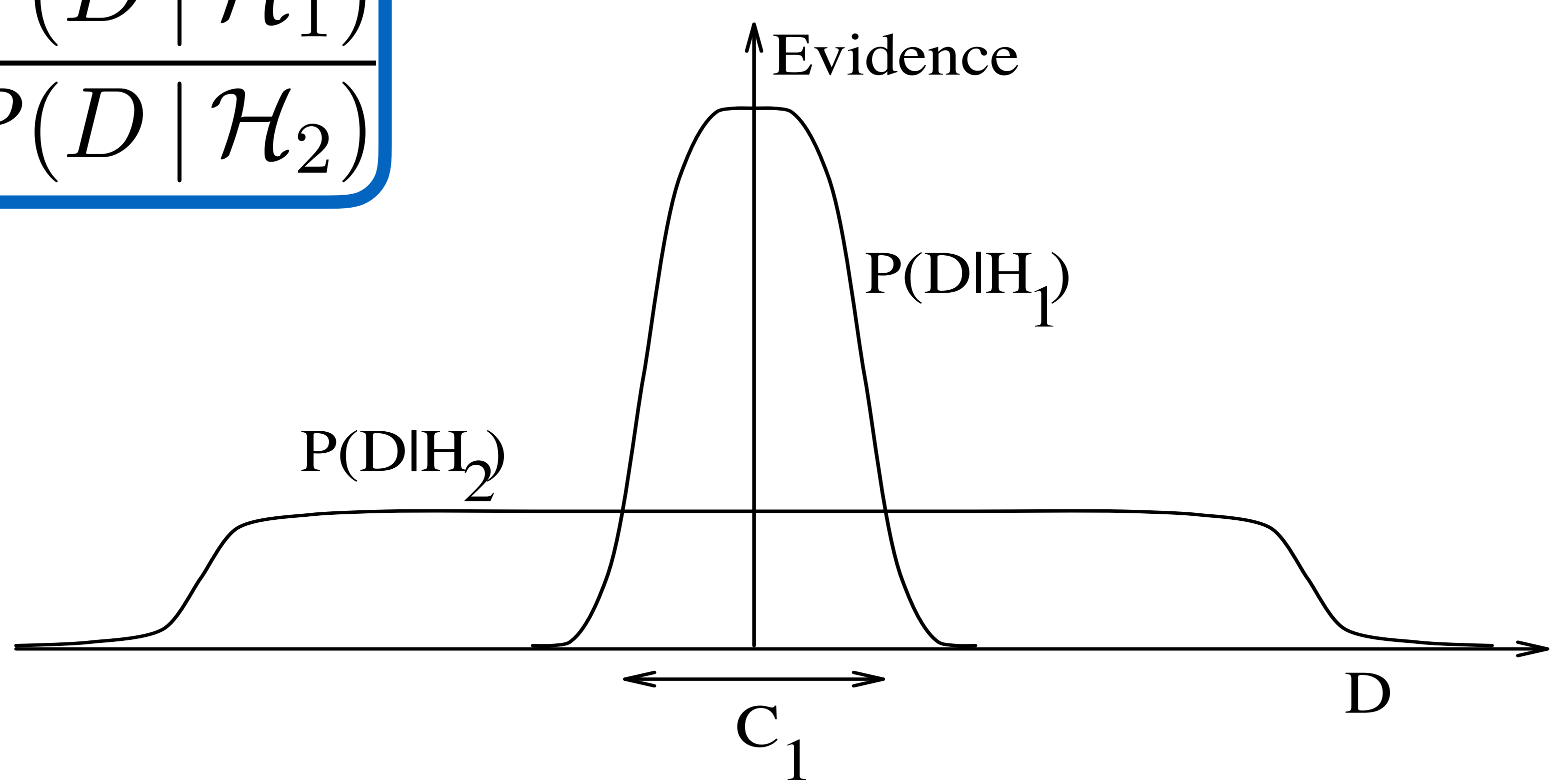
Dirac: porque es más *bella*
o: ¡porque esta estrategia viene funcionando bien!

Inferencia Bayesiana: lo dice la cuenta

Bayes Factor

$$\frac{P(\mathcal{H}_1\,|\,D)}{P(\mathcal{H}_2\,|\,D)} = \frac{P(\mathcal{H}_1)}{P(\mathcal{H}_2)}\boxed{\frac{P(D\,|\,\mathcal{H}_1)}{P(D\,|\,\mathcal{H}_2)}}$$

Caricatura de modelos sencillos



-1, 3, 7, 11, ...

¿Cuáles son los próximos dos números de la secuencia?

15, 19

sumar 4 al anterior

-19.9, 1043.8

evaluar sobre el anterior *x:*

$$-x^3/11 + 9/11x^2 + 23/11$$

Formalizando:

$\mathcal{H}_a$ progresión aritmética, sumar *n*

$\mathcal{H}_c$ función cúbica a partir del anterior $x \to cx^3 + dx^2 + e$
(con *c, d, e:* fracciones)

¡¡Tomemos *priors* iguales!!

Calculamos la *Evidencia* para cada modelo

$\mathcal{H}_a$ progresión aritmética, sumar *n*

$\mathcal{H}_c$ función cúbica a partir del anterior $x \rightarrow cx^3 + dx^2 + e$

(con *c, d, e:* fracciones)

(Tomamos intervalos -50 a 50)

$$P(D \,|\, \mathcal{H}_a) = \frac{1}{101}\frac{1}{101} = 0.00010$$

dónde empiezo y cuánto salto

$$\begin{aligned} P(D \,|\, \mathcal{H}_c) &= \left(\frac{1}{101}\right) \underbrace{\left(\frac{4}{101}\frac{1}{50}\right)}_{c} \underbrace{\left(\frac{4}{101}\frac{1}{50}\right)}_{d} \underbrace{\left(\frac{2}{101}\frac{1}{50}\right)}_{e} \\ &= 0.0000000000025 = 2.5 \times 10^{-12}. \end{aligned}$$

Los *odds* son de 40 millones a 1..

Lo mismo pasa en la ciencia: Copérnico vs. *epiciclos*

Inferencia en dos niveles

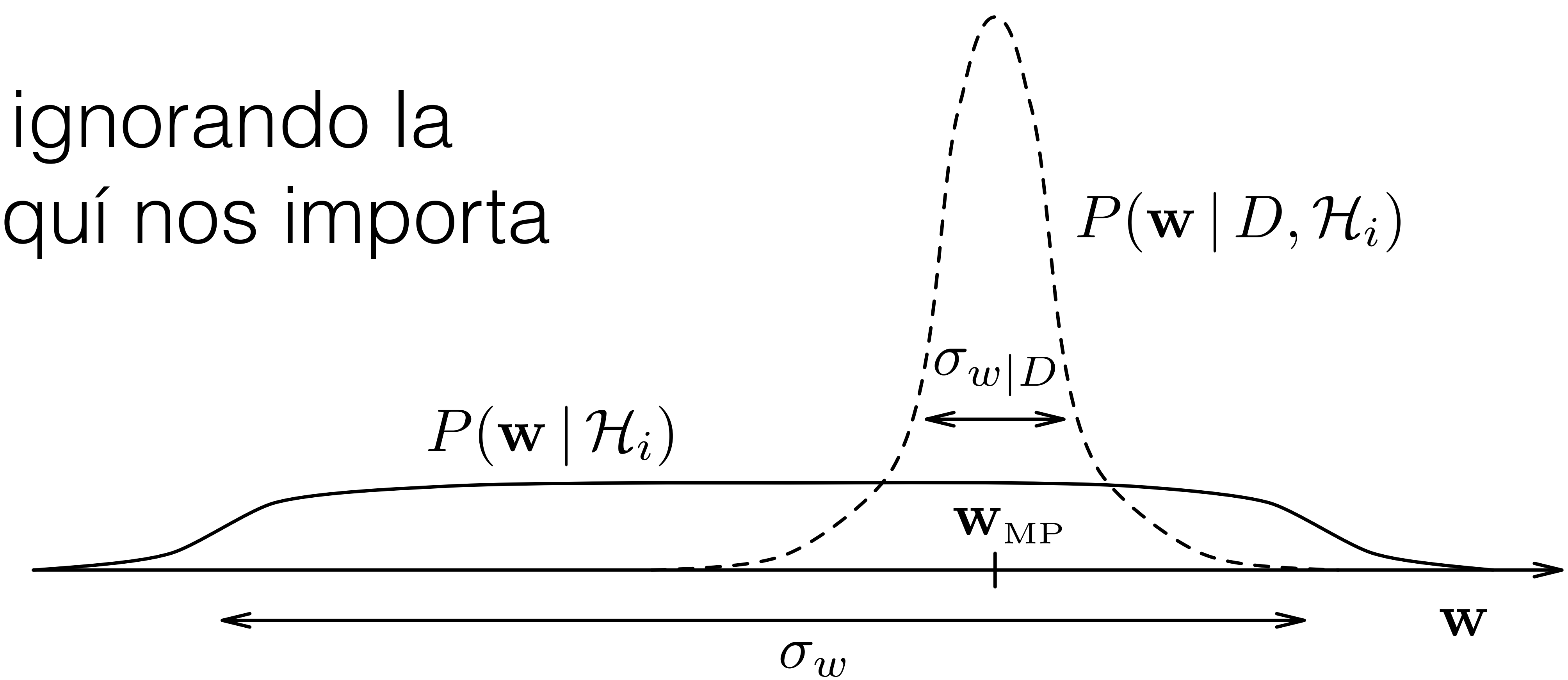
1) Ajuste de Modelos

$$P(\mathbf{w} \mid D, \mathcal{H}_i) = \frac{P(D \mid \mathbf{w}, \mathcal{H}_i) P(\mathbf{w} \mid \mathcal{H}_i)}{P(D \mid \mathcal{H}_i)}$$

2) Comparación de Modelos

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

Veníamos ignorando la evidencia.. aquí nos importa



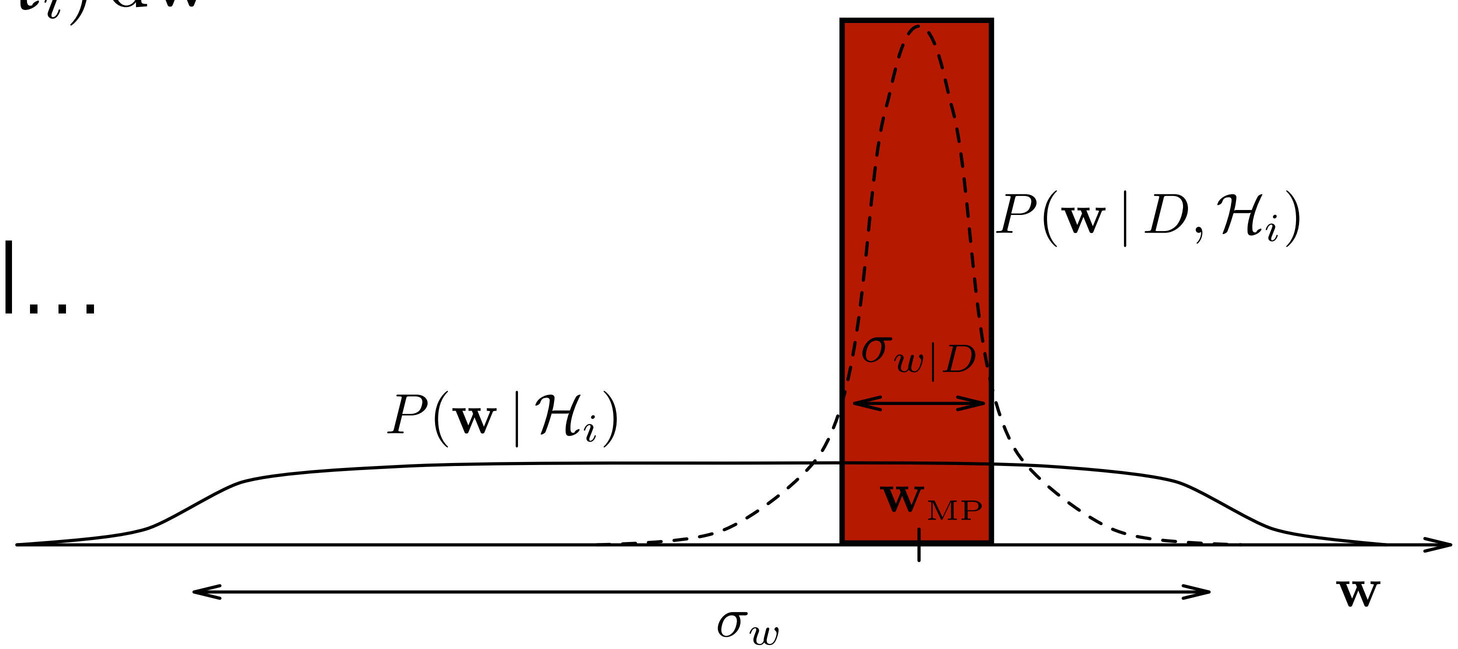
2) Comparación de Modelos

$$P(\mathcal{H}_i \mid D) \propto P(D \mid \mathcal{H}_i) P(\mathcal{H}_i)$$

Evidencia
(era la normalización)

$$P(D \mid \mathcal{H}_i) = \int P(D \mid \mathbf{w}, \mathcal{H}_i) P(\mathbf{w} \mid \mathcal{H}_i) \, \mathrm{d}\mathbf{w}$$

Aproximamos la intergral...



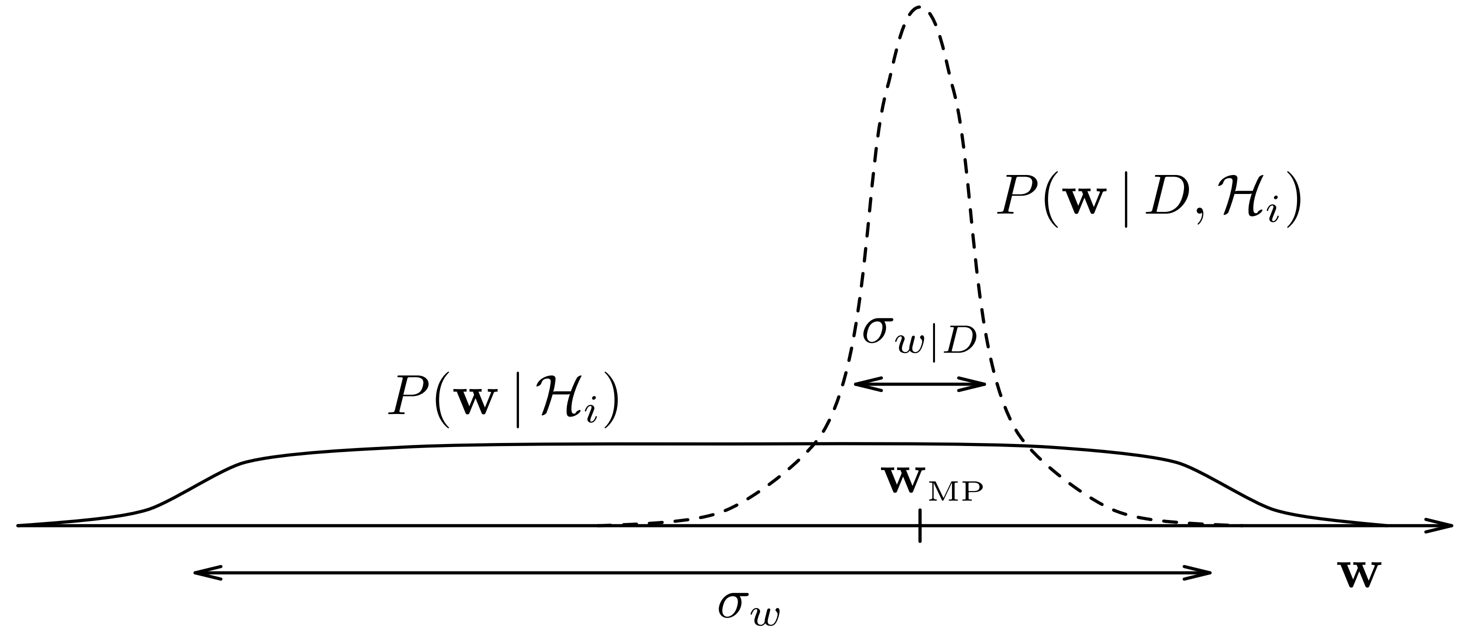
$$P(D \mid \mathcal{H}_i) \simeq \underbrace{P(D \mid \mathbf{w}_{\mathrm{MP}}, \mathcal{H}_i)} \times \underbrace{P(\mathbf{w}_{\mathrm{MP}} \mid \mathcal{H}_i) \, \sigma_{w|D}}$$

$$\text{Evidence} \simeq \text{Best fit likelihood} \times \text{Occam factor}$$

El factor de Occam

prior plana:

$$P(\mathbf{w}_{\mathrm{MP}} \,|\, \mathcal{H}_i) = 1/\sigma_w$$



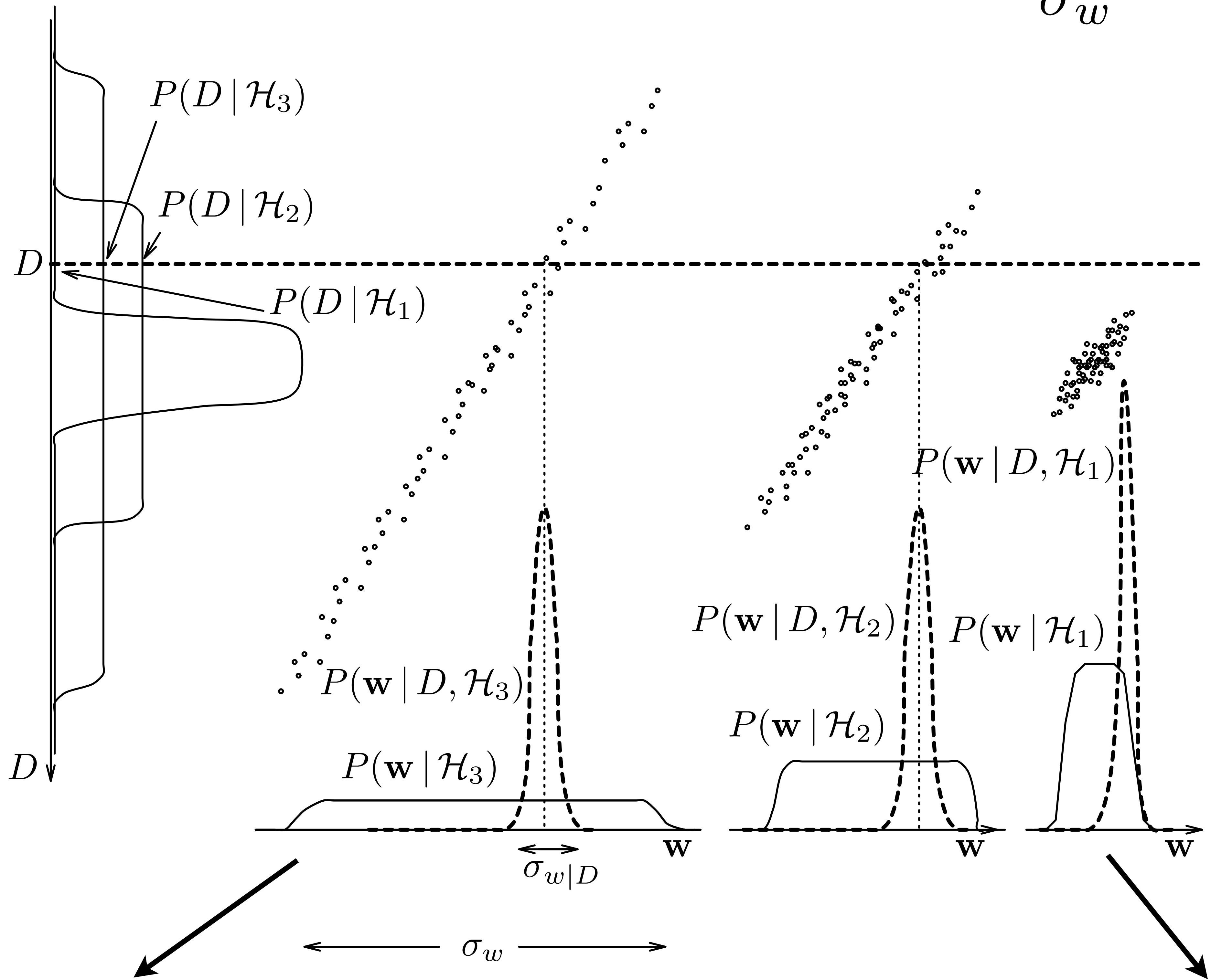
$$\begin{array}{rcl} P(D \,|\, \mathcal{H}_i) & \simeq & \underbrace{P(D \,|\, \mathbf{w}_{\mathrm{MP}}, \mathcal{H}_i)} \times \underbrace{P(\mathbf{w}_{\mathrm{MP}} \,|\, \mathcal{H}_i)\, \sigma_{w|D}} \\ \text{Evidence} & \simeq & \text{Best fit likelihood} \times \text{Occam factor} \end{array}$$

$$\text{Factor de Occam} = \frac{\sigma_{w|D}}{\sigma_w}$$

-penaliza modelos con rango grande de parámetros
-penaliza modelos cuyos parámetros requieren un ajuste fino a los datos

...¡y todo esto elevado al número de parámetros!

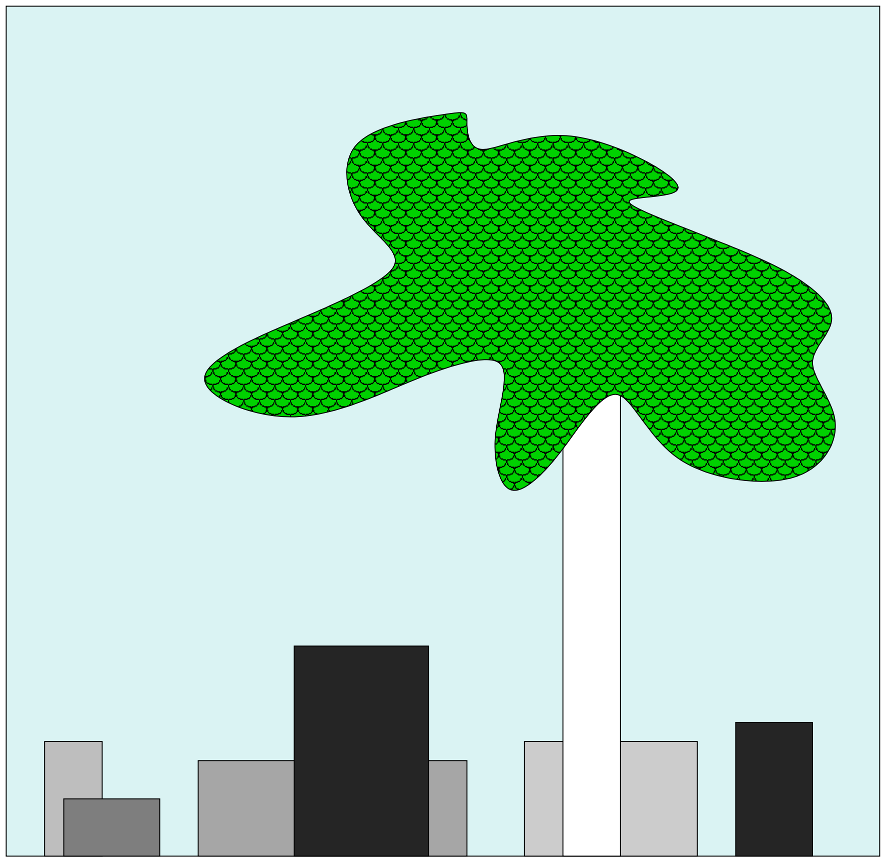
$$P(D|\mathcal{H}_i) \simeq P(D|\mathbf{w}_{MP}, \mathcal{H}_i) \times \frac{\sigma_{w|D}}{\sigma_w}$$

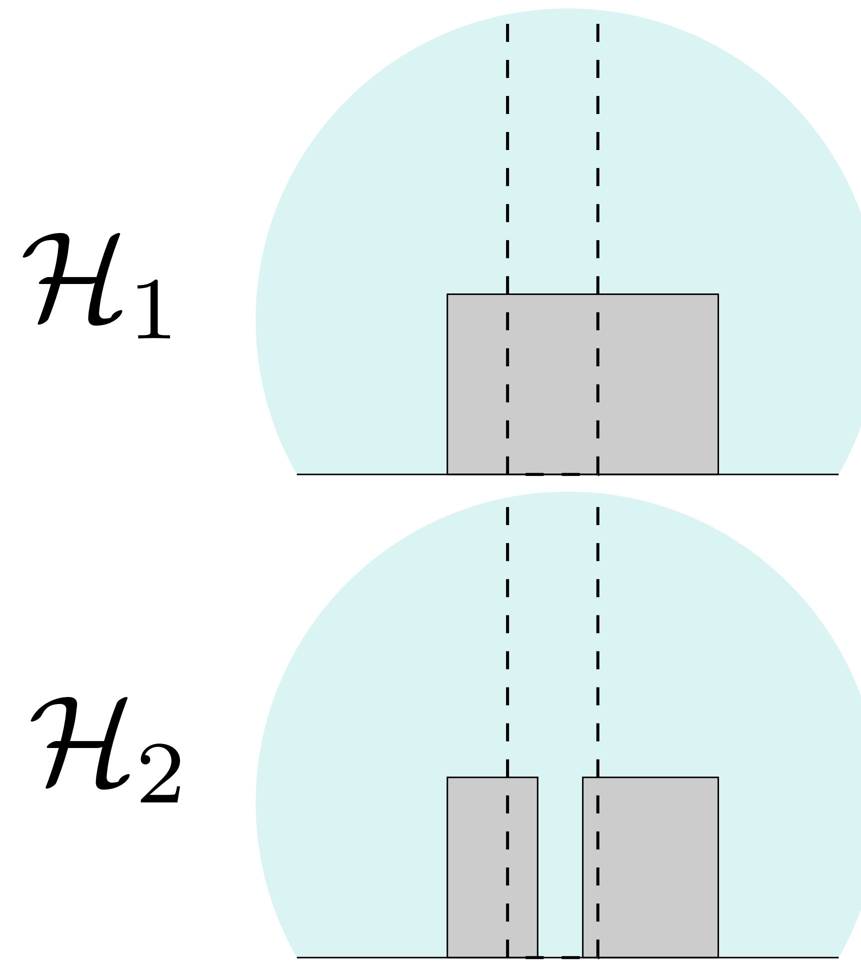


Buen *likelihood*, mal *Occam*

Buen *Occam*, mal *likelihood*

De vuelta al árbol...





$$P(D\,|\,\mathcal{H}_1) = \frac{1}{20}\frac{1}{20}\frac{1}{20}\frac{1}{16}$$

$$P(D\,|\,\mathcal{H}_2) \simeq \frac{1}{20}\frac{1}{20}\frac{10}{20}\frac{1}{16}\frac{1}{20}\frac{1}{20}\frac{10}{20}\frac{1}{16}\frac{2}{2}$$

(aproximando restricciones en los parámetros)

$$\frac{P(D\,|\,\mathcal{H}_1)P(\mathcal{H}_1)}{P(D\,|\,\mathcal{H}_2)P(\mathcal{H}_2)} = \frac{1}{\frac{1}{20}\frac{10}{20}\frac{10}{20}\frac{1}{16}} \simeq 1000/1$$

factores de Occam

Frecuentistas vs. Bayesianos

$$aaabaaaabaab$$

$$P(r \leq 3 \,|\, n = 12, \mathcal{H}_0) \;=\; \sum_{r=0}^{3} \binom{n}{r} {}^1\!/{}_2{}^n = \left(\tbinom{12}{0} + \tbinom{12}{1} + \tbinom{12}{2} + \tbinom{12}{3}\right) {}^1\!/{}_2{}^{12} = \;0.07$$

pero…

$$P(n \,|\, \mathcal{H}_0, r) = \binom{n-1}{r-1} {}^1\!/{}_2{}^n$$

$$P(n \geq 12 \,|\, r = 3, \mathcal{H}_0) = 0.03$$

Práctica

Wagenmakers & Lee: 5.3