ESTADÍSTICA BAYESIANA

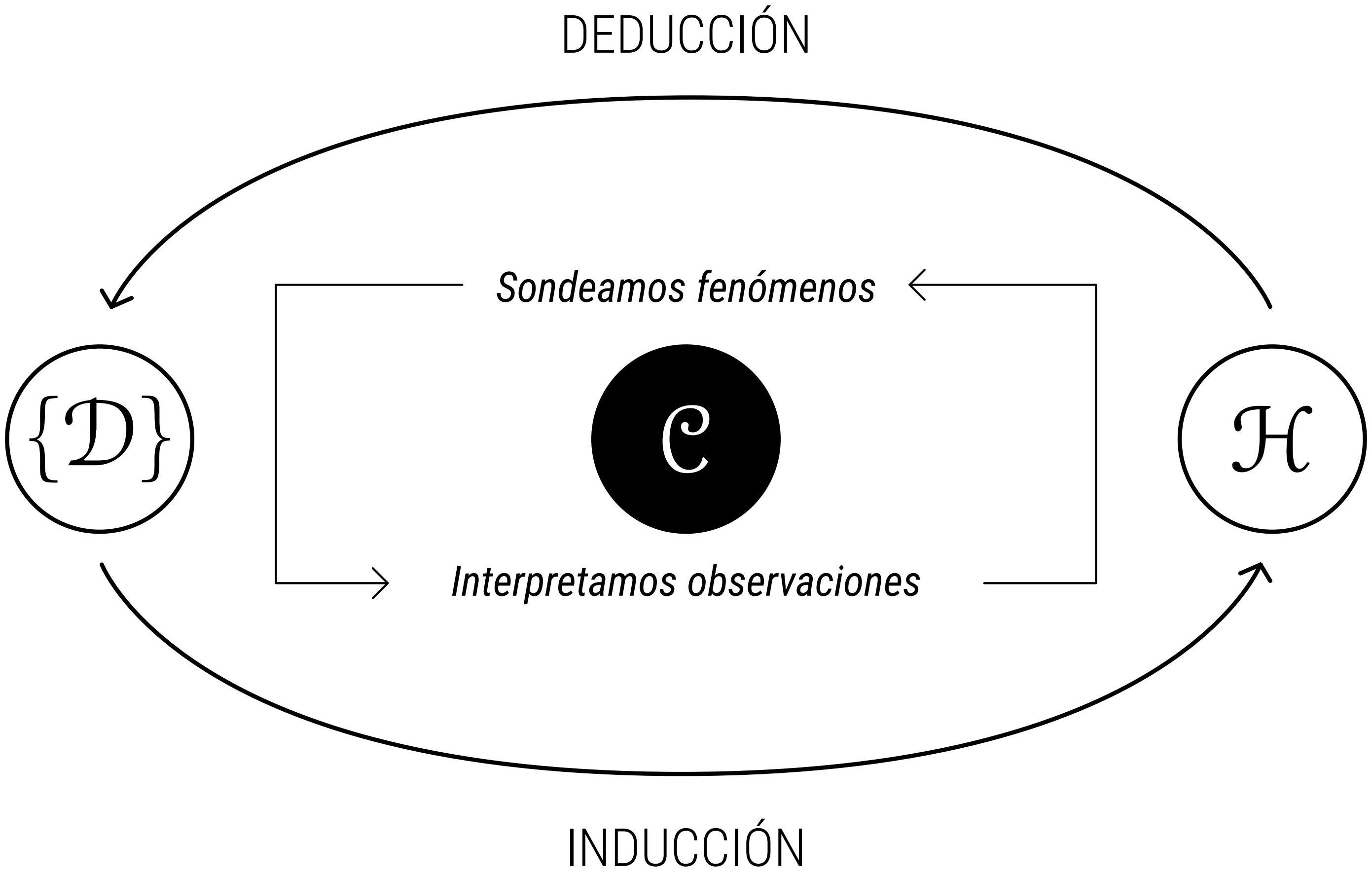
UN CONJUNTO DE PROCEDIMIENTOS ÚTILES

Por
Alfredo J. Mejía

ESTRUCTURA

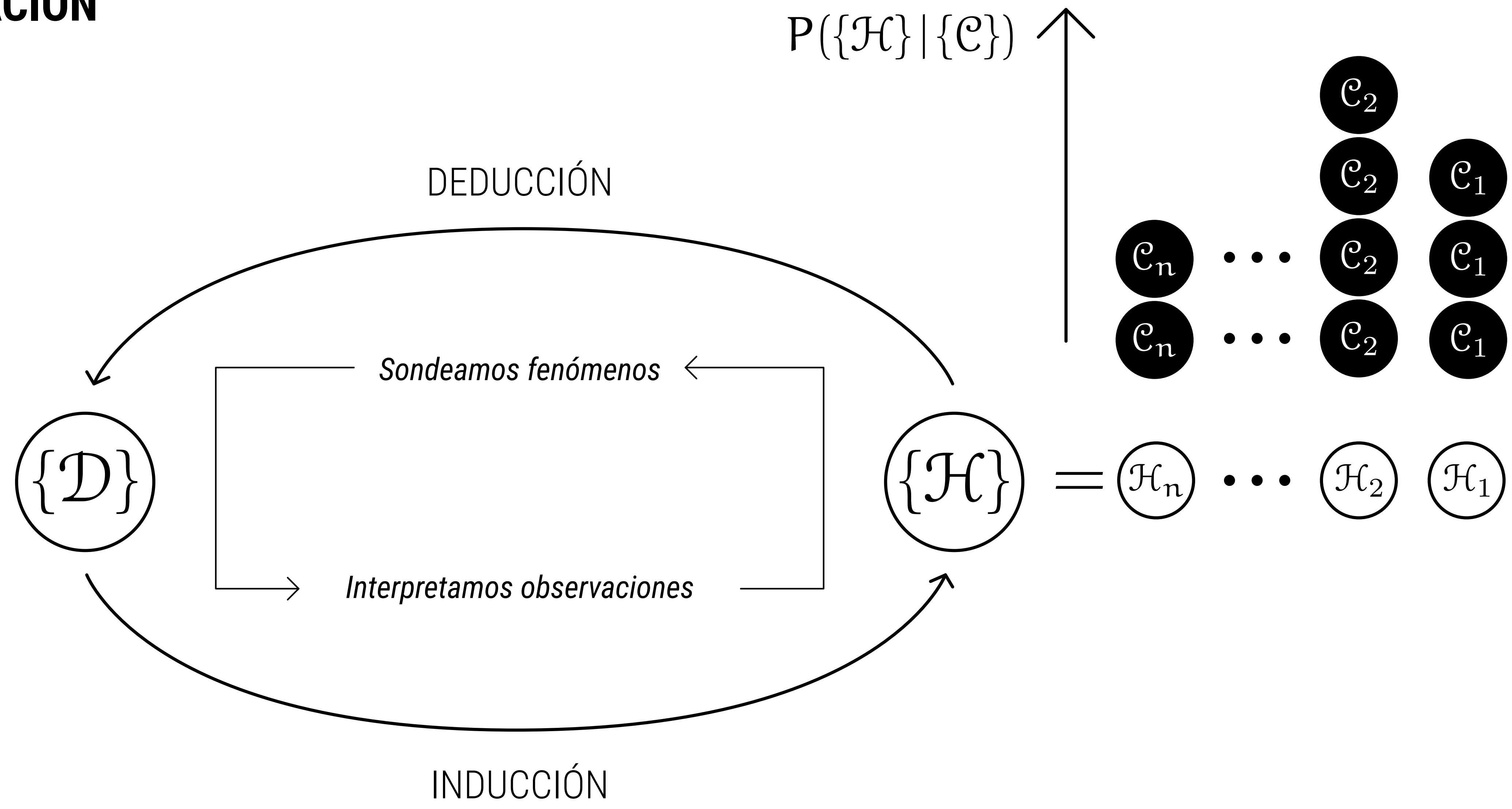
- Motivación
- Probabilidades 101
- Teorema de Bayes
- Regresión Lineal
  - Máxima Verosimilitud
  - Marginalización
  - El rol de la Distribución Previa
  - ¿Cómo asignar la Distribución Previa?
- Resumen

MOTIVACIÓN



(Sivia+2006)

MOTIVACIÓN

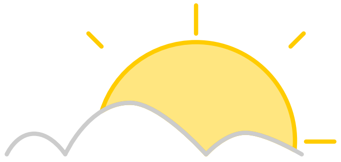


(Sivia+2006)

PROBABILIDADES 101

PROBABILIDAD: grado de certidumbre sobre la ocurrencia de un evento.

$$\mathrm{L} \qquad \mathrm{P}(\mathrm{L}\,|\,\mathcal{C}) = 0,76$$



$$\mathrm{S} = \bar{\mathrm{L}} \qquad \mathrm{P}(\mathrm{S}\,|\,\mathcal{C}) = 0,24$$

REGLAS DE COX:

$$\sum_i \mathrm{P}(\mathcal{H}_i\,|\,\mathcal{C}) = 1$$

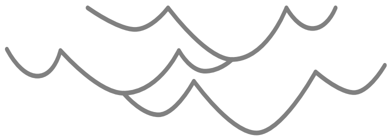
$$\mathrm{P}(\mathcal{H}_A, \mathcal{H}_B\,|\,\mathcal{C}) = \mathrm{P}(\mathcal{H}_A\,|\,\mathcal{H}_B, \mathcal{C}) \times \mathrm{P}(\mathcal{H}_B\,|\,\mathcal{C})$$

$$\mathrm{P}(\mathrm{S}\,|\,\mathcal{C}) = 1 - \mathrm{P}(\mathrm{L}\,|\,\mathcal{C})$$

La certidumbre de ocurrencia de un evento define también su incertidumbre.

PROBABILIDADES 101

PROBABILIDAD: grado de certidumbre sobre la ocurrencia de un evento.



N $\quad P(N \,|\, \mathcal{C}) = 0,63$

L $\quad P(L \,|\, N, \mathcal{C}) = 0,81$

REGLAS DE COX:

$$\sum_i P(\mathcal{H}_i \,|\, \mathcal{C}) = 1$$

$$P(\mathcal{H}_A, \mathcal{H}_B \,|\, \mathcal{C}) = P(\mathcal{H}_A \,|\, \mathcal{H}_B, \mathcal{C}) \times P(\mathcal{H}_B \,|\, \mathcal{C})$$

$$P(L, \bar{N} \,|\, \mathcal{C}) \leqslant [1 - P(L \,|\, N, \mathcal{C})] \times P(\bar{N} \,|\, \mathcal{C})$$

$$P(L, \bar{N} \,|\, \mathcal{C}) \leqslant 0,07$$

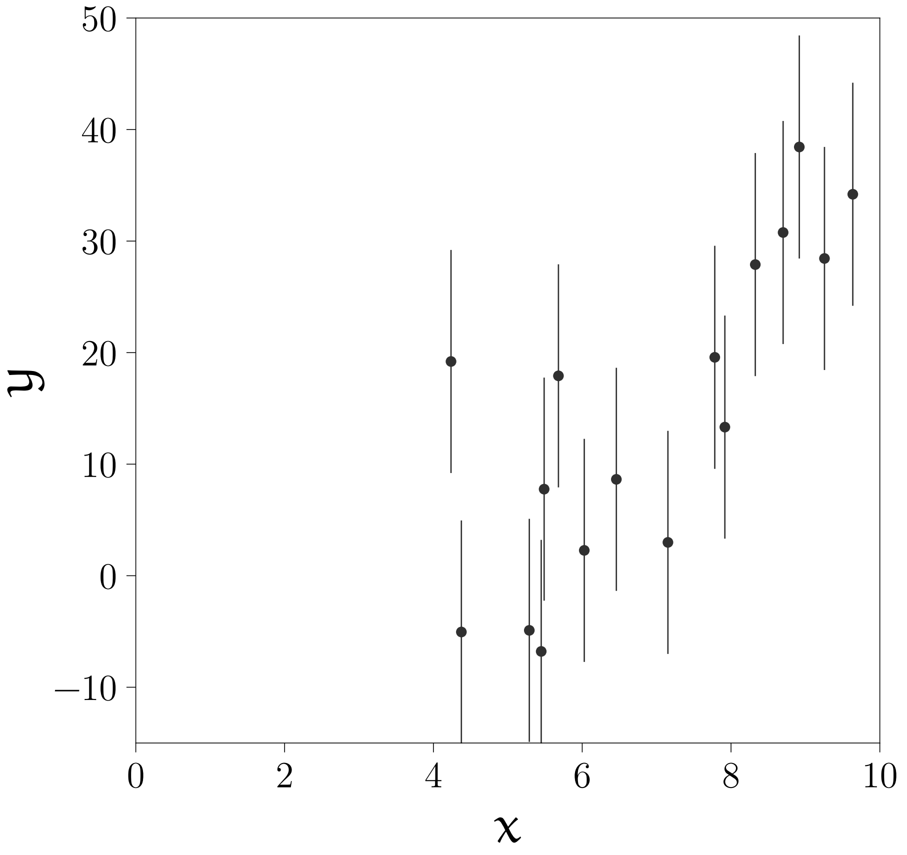
TEOREMA DE BAYES

$$\overbrace{P(\mathcal{H} \mid \mathcal{D}, \mathcal{C})}^{\text{Estado de conocimiento sobre la hipótesis posterior a la observación}} = \frac{\overbrace{P(\mathcal{D} \mid \mathcal{H}, \mathcal{C})}^{\text{Probabilidad de la observación dada la hipótesis}} \times \overbrace{P(\mathcal{H} \mid \mathcal{C})}^{\text{Estado de conocimiento sobre la hipótesis previo a la observación}}}{\underbrace{P(\mathcal{D} \mid \mathcal{C})}_{\text{Evidencia o distribución posterior marginalizada}}}$$

TEOREMA DE BAYES

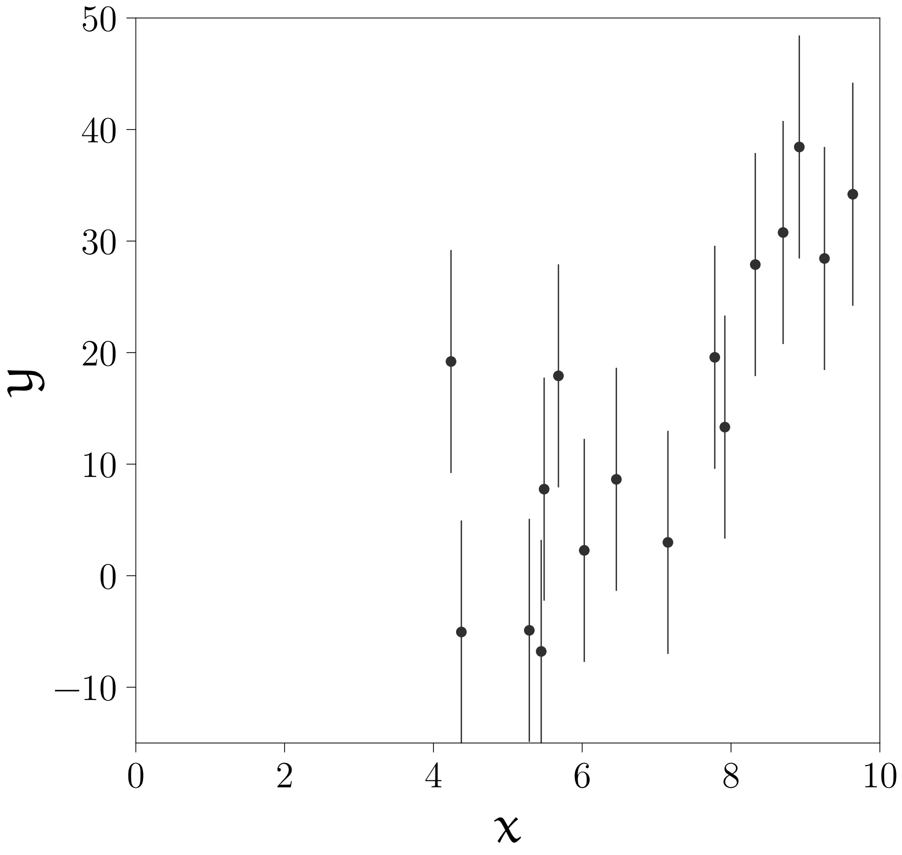
$$\overbrace{P(\mathcal{H}\mid\mathcal{D},\mathcal{C})}^{\substack{\text{Estado de conocimiento}\\ \text{sobre la hipótesis}\\ \text{posterior a la observación}}} = \frac{\overbrace{\mathcal{L}(\mathcal{H}\mid\mathcal{D},\mathcal{C})}^{\substack{\text{Verosimilitud de la hipótesis}\\ \text{frente a la observación}}} \times \overbrace{P(\mathcal{H}\mid\mathcal{C})}^{\substack{\text{Estado de conocimiento}\\ \text{sobre la hipótesis}\\ \text{previo a la observación}}}}{\underbrace{P(\mathcal{D}\mid\mathcal{C})}_{\substack{\text{Evidencia o distribución}\\ \text{posterior marginalizada}}}}$$

VEROSIMILITUD — REGRESIÓN LINEAL



$$\overbrace{\mathcal{L}(\overbrace{m, b}^{\mathcal{H}} \mid \overbrace{\{y\}}^{\mathcal{D}}, \overbrace{\{x\}, \{\sigma_y\}}^{\mathcal{C}})}^{\text{Verosimilitud}}$$

VEROSIMILITUD — REGRESIÓN LINEAL

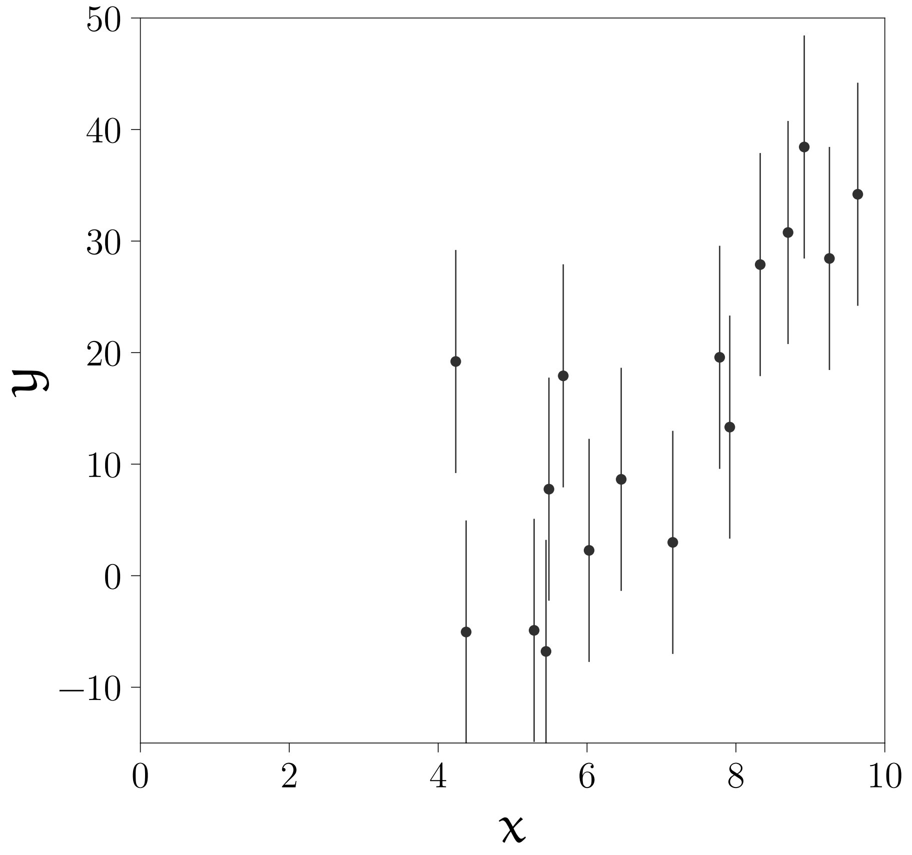


$$\overbrace{\mathcal{L}(\overbrace{m,b}^{\mathcal{H}}\mid\overbrace{\{y\}}^{\mathcal{D}},\overbrace{\{x\},\{\sigma_y\}}^{\mathcal{C}})}^{\text{Verosimilitud}}$$

$$\{y\}\sim\text{Gaussiana}(\mu(\{x\}),\{\sigma_y\})$$

$$\mu(x)\equiv m\,x+b$$

VEROSIMILITUD — REGRESIÓN LINEAL



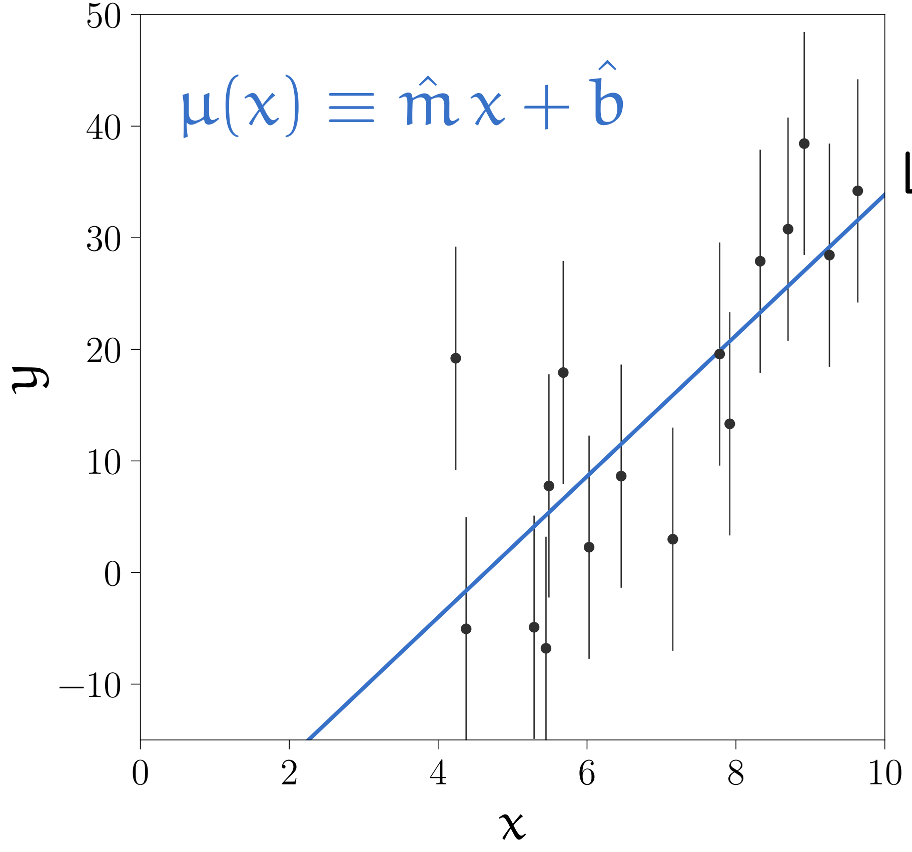
Verosimilitud

$$\mathcal{L}(\underbrace{m, b}_{\mathcal{H}} \mid \underbrace{\{y\}}_{\mathcal{D}}, \underbrace{\{x\}, \{\sigma_y\}}_{\mathcal{C}})$$

$$\{y\} \sim \text{Gaussiana}(\mu(\{x\}), \{\sigma_y\}) \qquad \mu(x) \equiv m\,x + b$$

$$\mathcal{L}(m, b \mid \{y\}, \{x\}, \{\sigma_y\}) \propto \prod_i \exp\left[-\frac{(\mu(m, b; x_i) - y_i)^2}{2{\sigma_y}_i^2}\right]$$

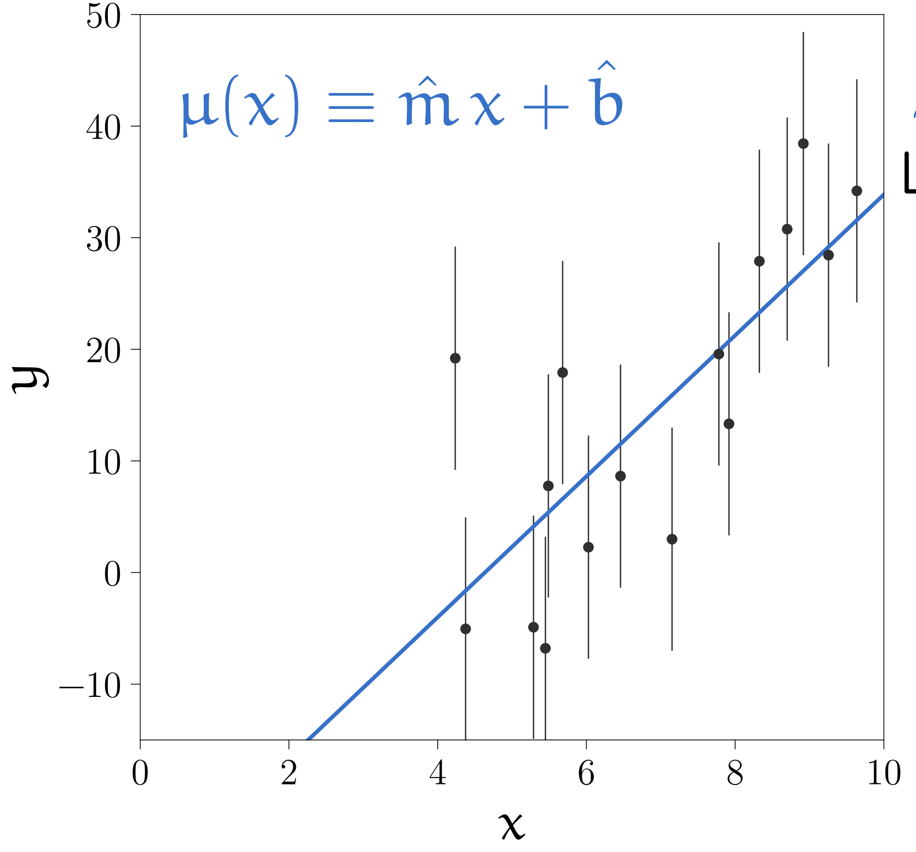
VEROSIMILITUD — REGRESIÓN LINEAL



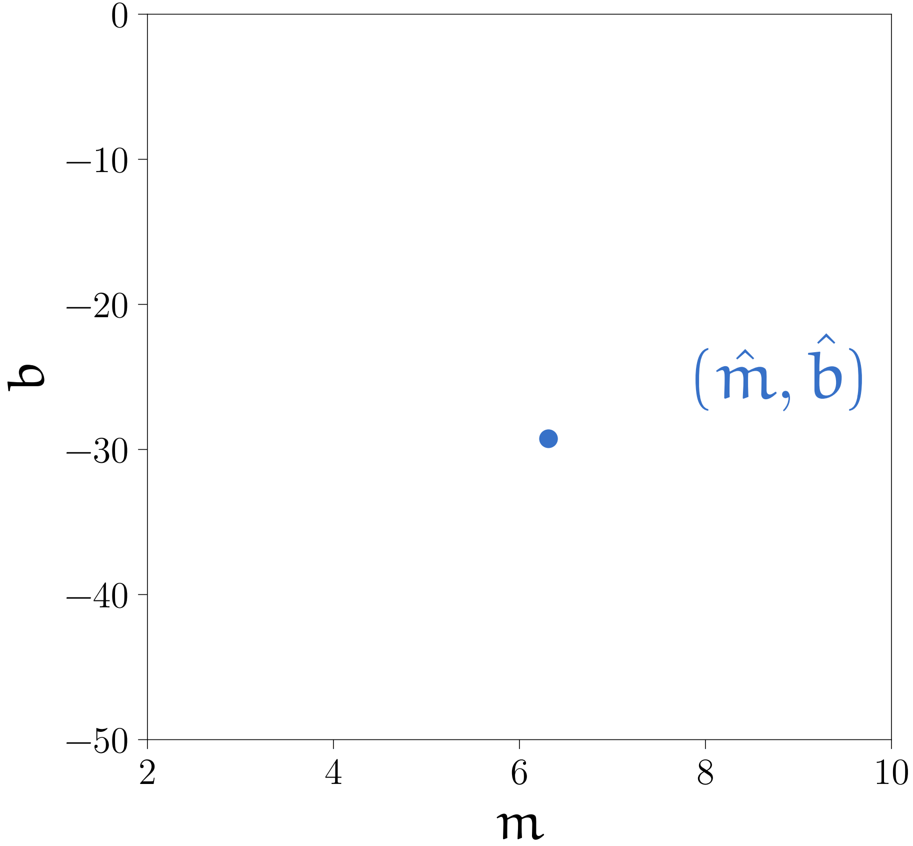
$$\mu(x) \equiv \hat{m}\,x + \hat{b}$$

$$\underbrace{\mathcal{L}_{\max} \equiv \mathcal{L}(\hat{m}, \hat{b} \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Máxima Verosimilitud}}$$

VEROSIMILITUD — REGRESIÓN LINEAL

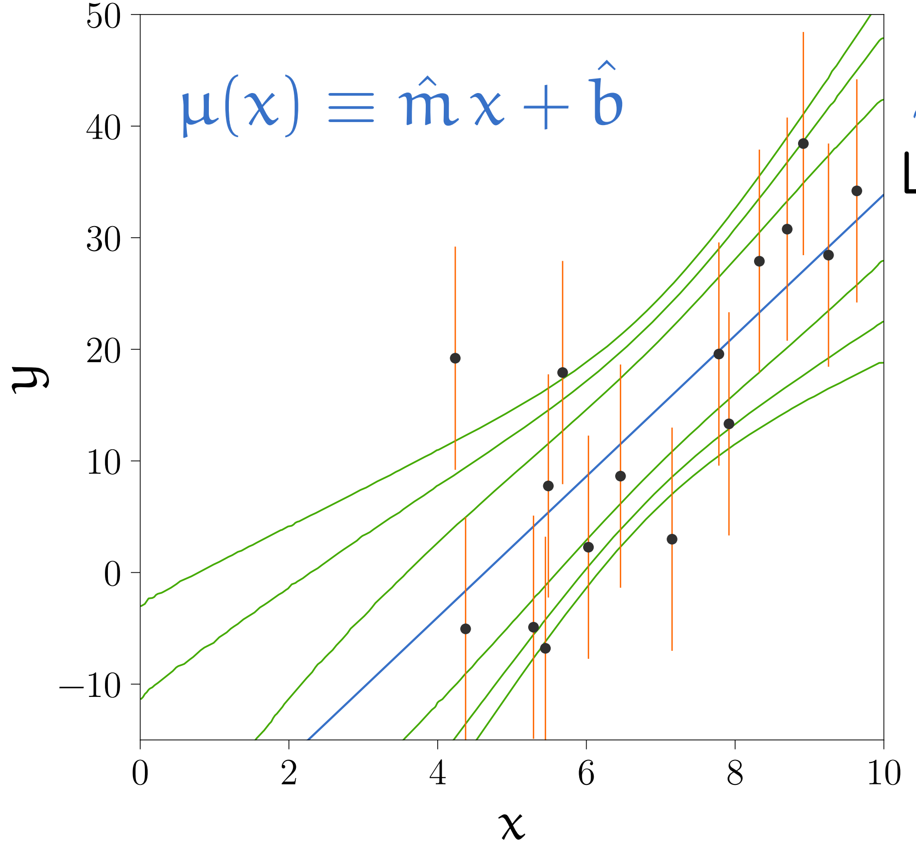


$$\underbrace{\mathcal{L}_{\max} \equiv \mathcal{L}(\hat{m}, \hat{b} \mid \{y\}, \{x\}, \{\sigma_y\})}_{\text{Máxima Verosimilitud}}$$

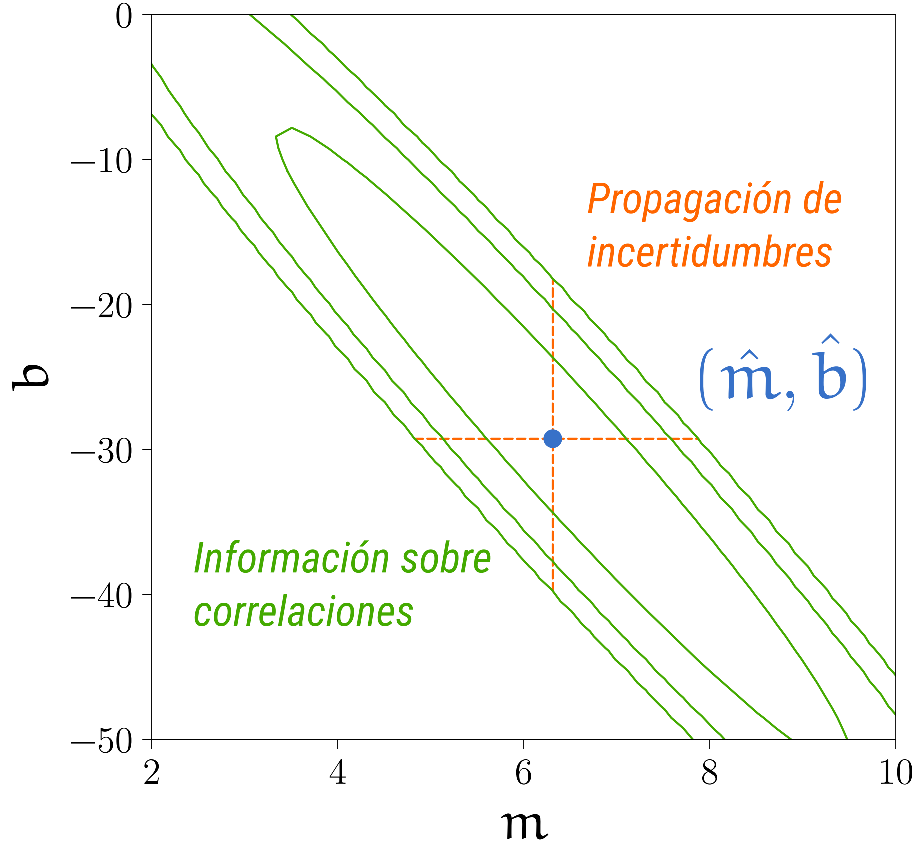


Un solo punto en el espacio de parámetros: ¡DECEPCIONANTE!

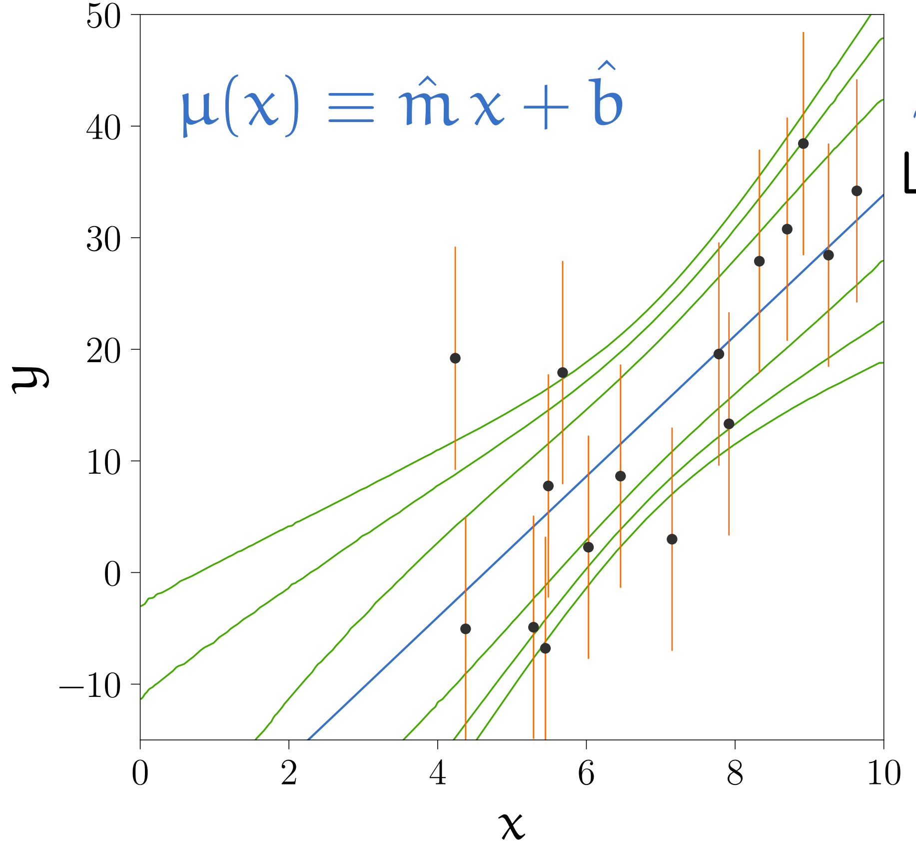
VEROSIMILITUD — REGRESIÓN LINEAL



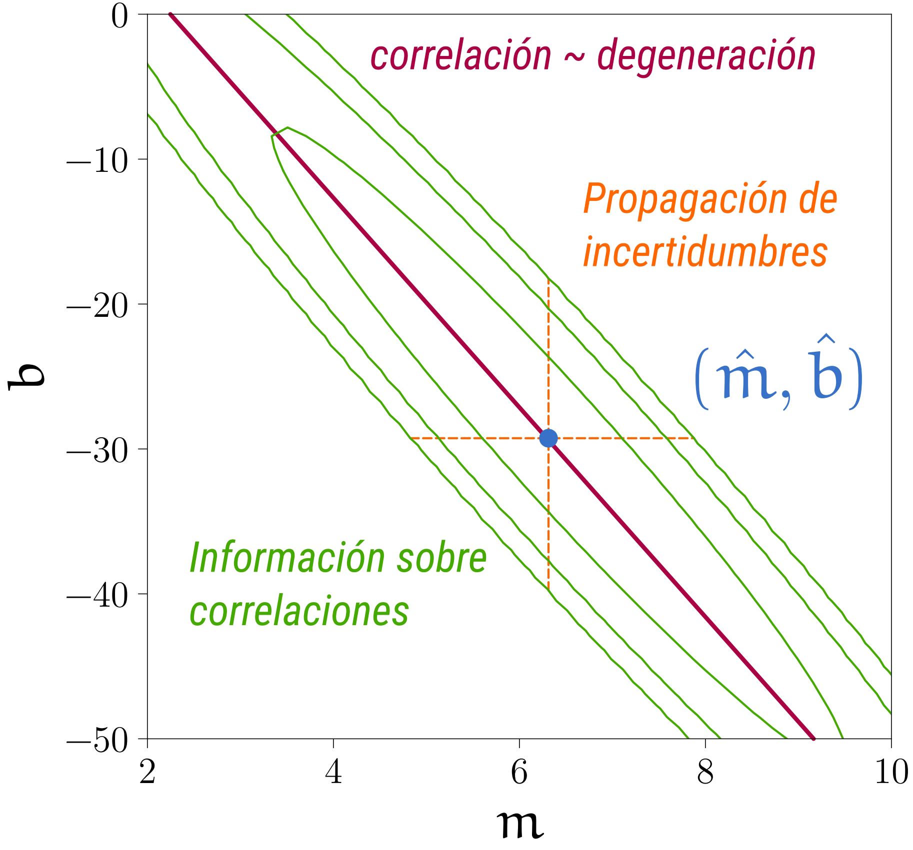
$$\underbrace{\mathcal{L}_{\max} \equiv \mathcal{L}(\hat{m}, \hat{b} \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Máxima Verosimilitud}} \quad \underbrace{\mathcal{L}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Verosimilitud}}$$



VEROSIMILITUD — REGRESIÓN LINEAL

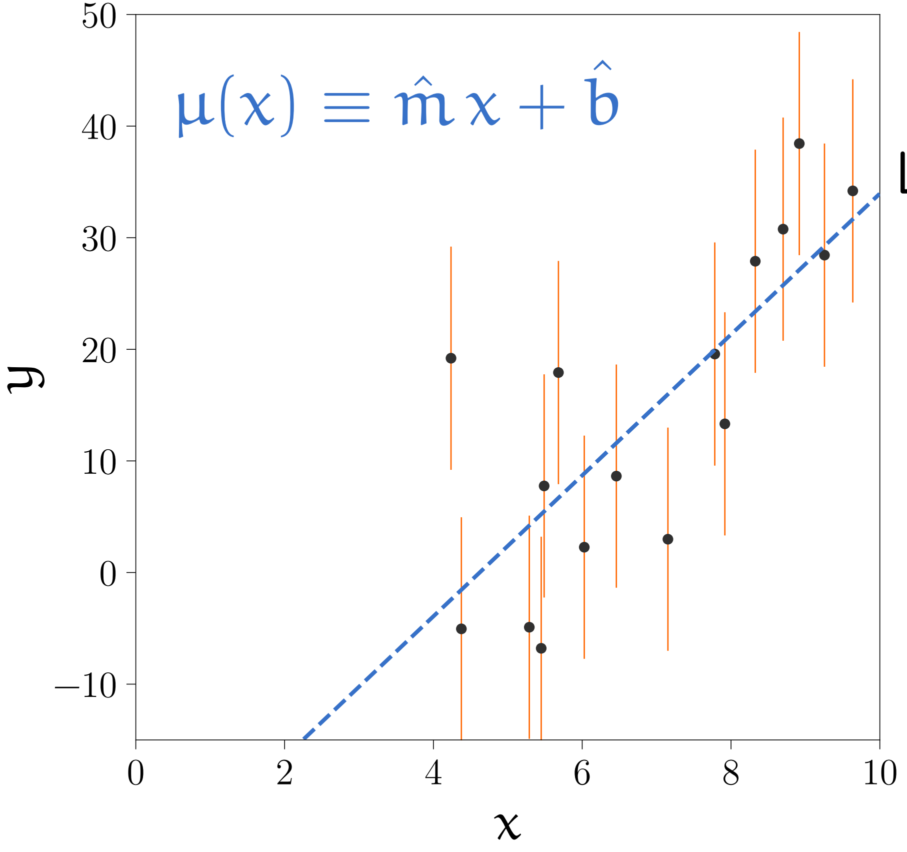


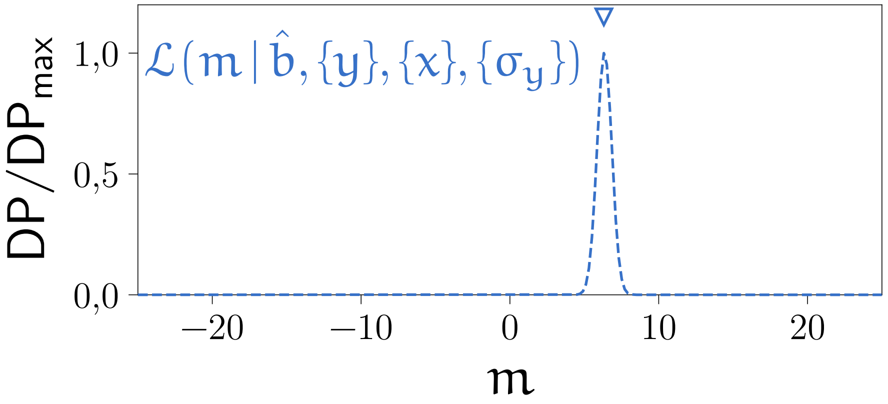
$$\underbrace{\mathcal{L}_{\max} \equiv \mathcal{L}(\hat{m}, \hat{b}\,|\,\{y\}, \{x\}, \{\sigma_y\})}_{\text{Máxima Verosimilitud}} \quad \underbrace{\mathcal{L}(m, b\,|\,\{y\}, \{x\}, \{\sigma_y\})}_{\text{Verosimilitud}}$$

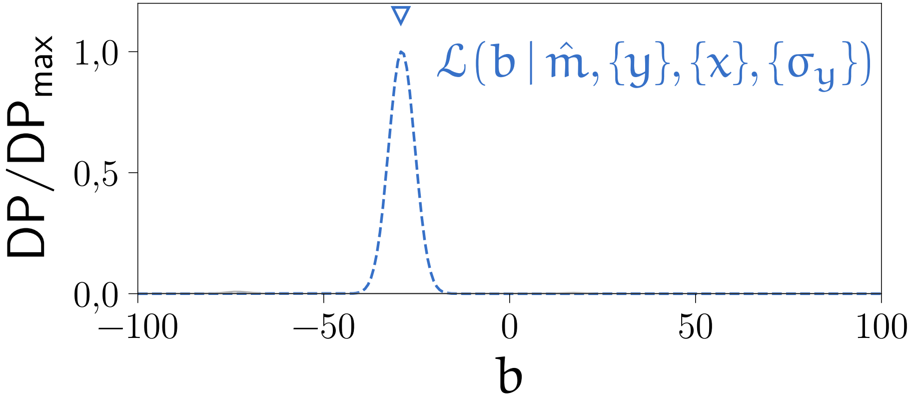


VEROSIMILITUD — REGRESIÓN LINEAL

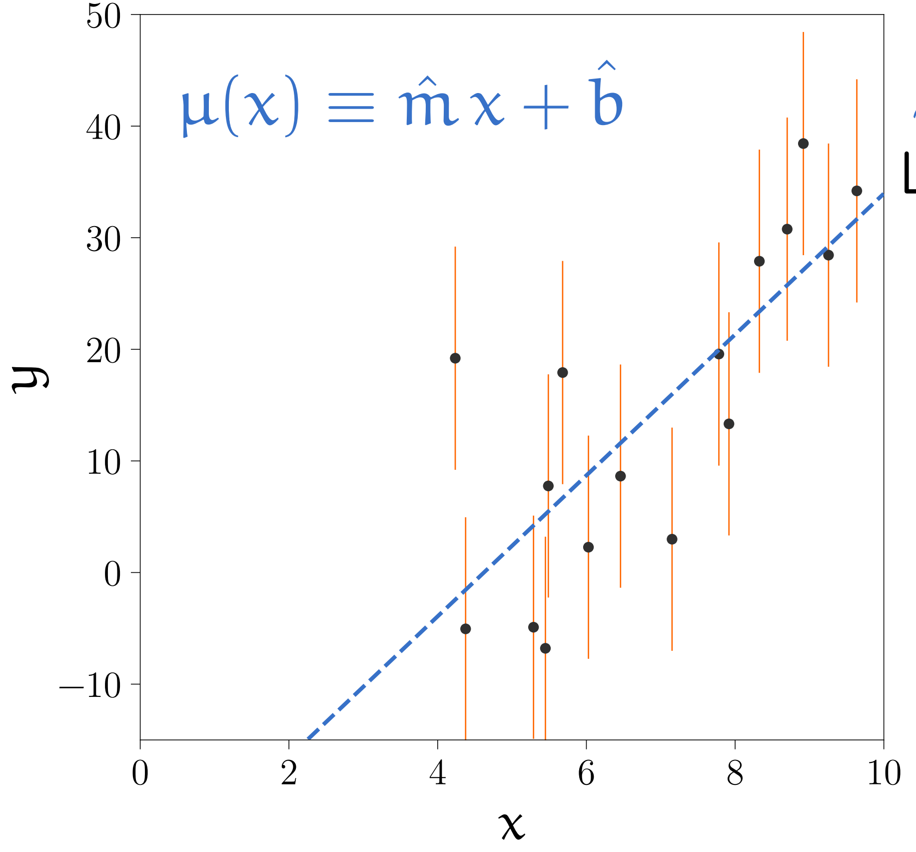
$$\underbrace{\mathcal{L}_{\max} \equiv \mathcal{L}(\hat{m}, \hat{b} \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Máxima Verosimilitud}}$$



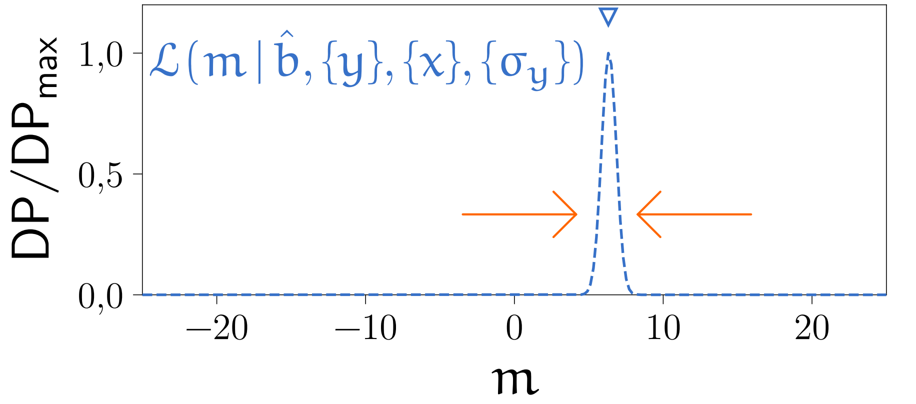


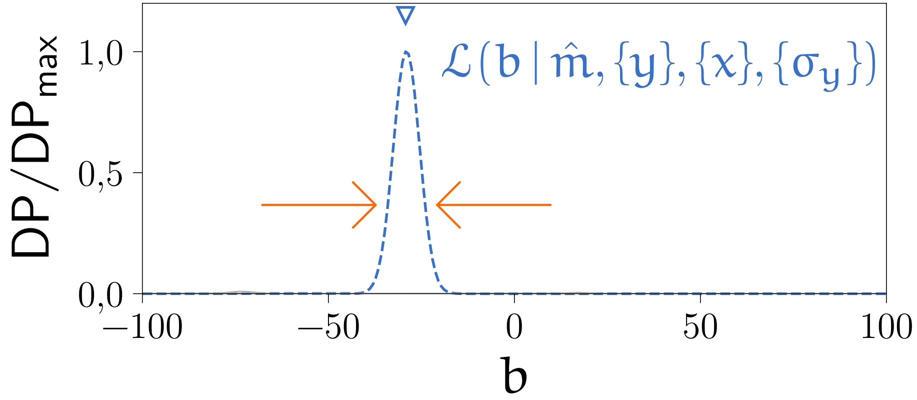


VEROSIMILITUD — REGRESIÓN LINEAL

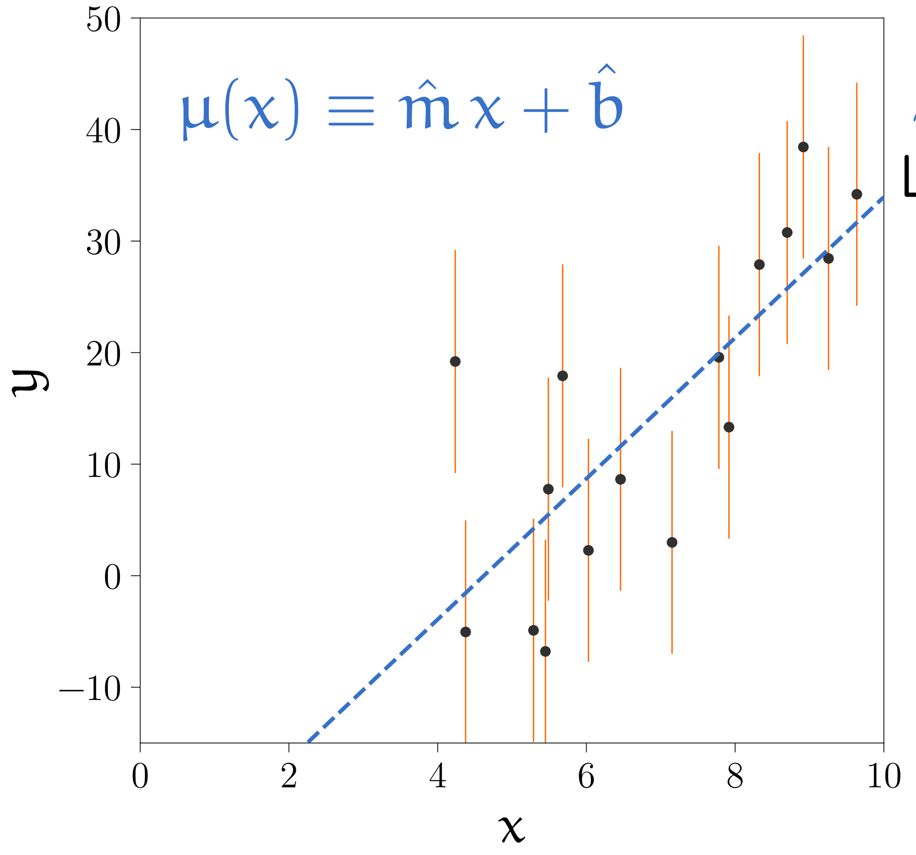


$$\underbrace{\mathcal{L}_{max} \equiv \mathcal{L}(\hat{m}, \hat{b} \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Máxima Verosimilitud}}$$

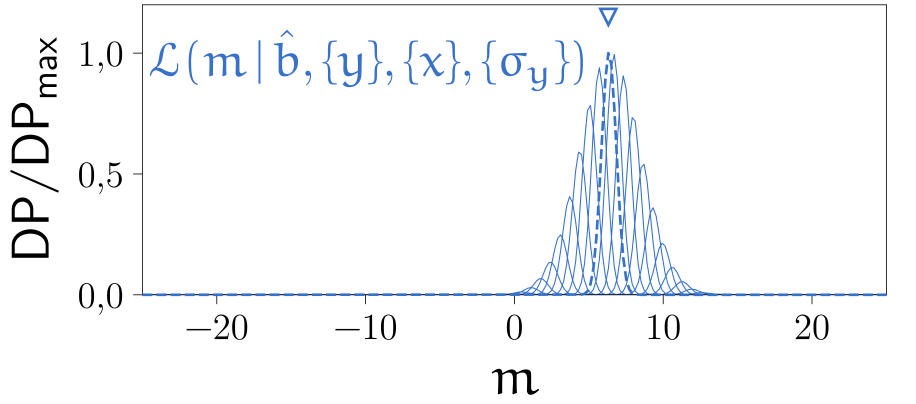


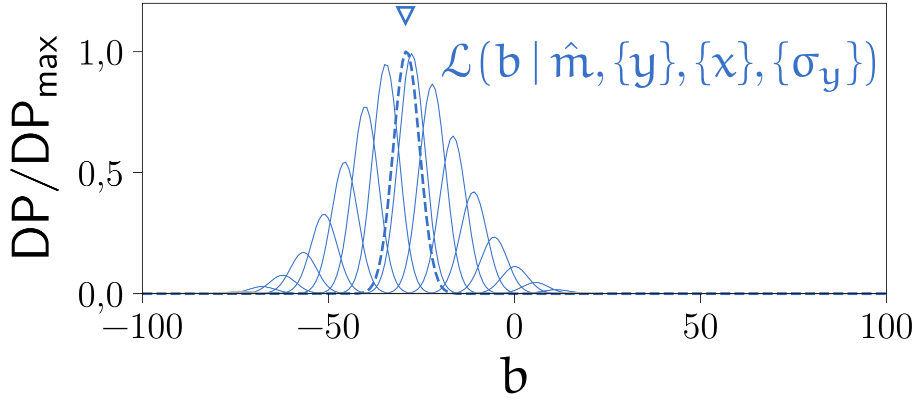


VEROSIMILITUD — REGRESIÓN LINEAL

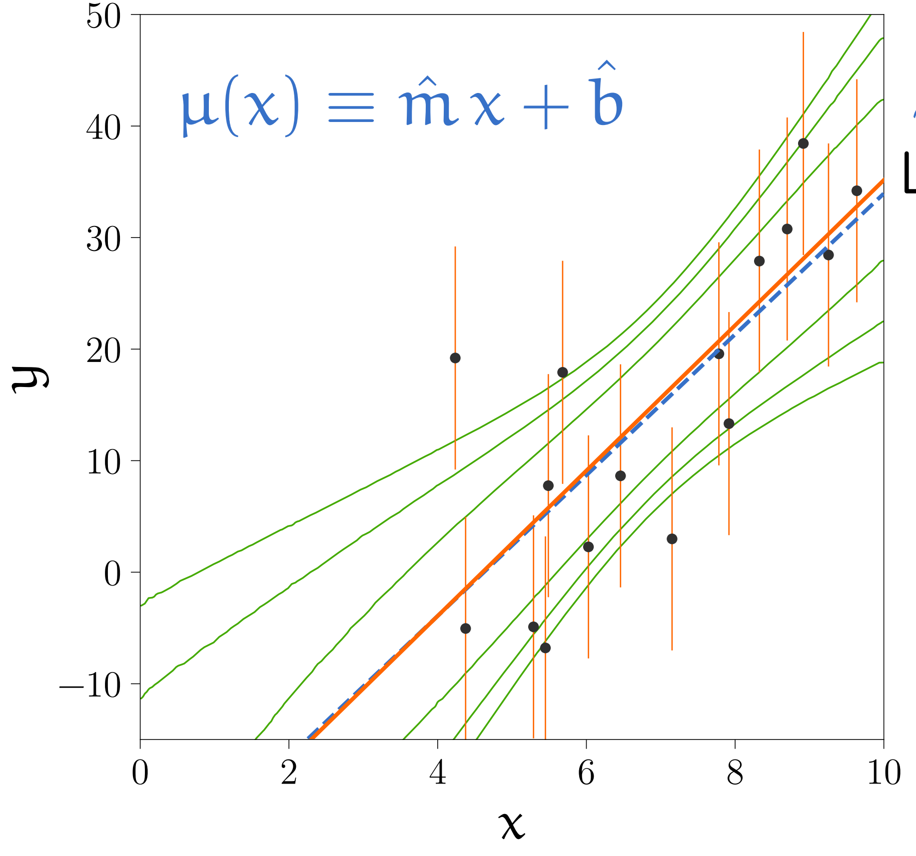


$$\mathcal{L}_{\max} \equiv \underbrace{\mathcal{L}(\hat{m}, \hat{b} \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Máxima Verosimilitud}}$$

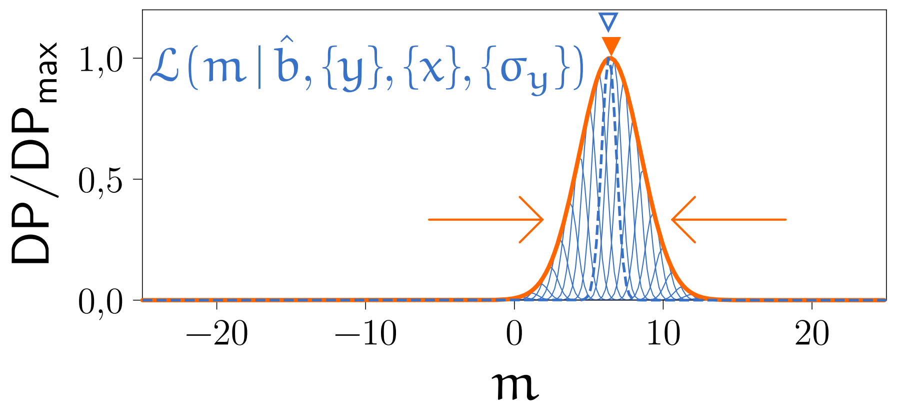




VEROSIMILITUD — REGRESIÓN LINEAL



$$\underbrace{\mathcal{L}_{max} \equiv \mathcal{L}(\hat{m}, \hat{b} \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Máxima Verosimilitud}}$$

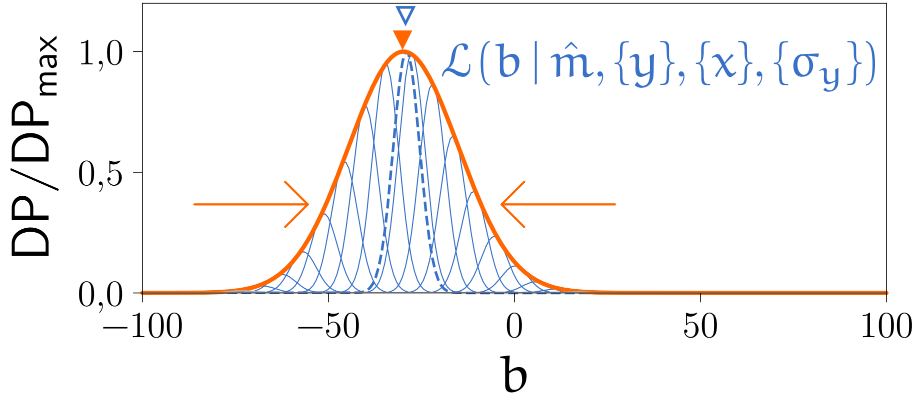


Verosimilitud

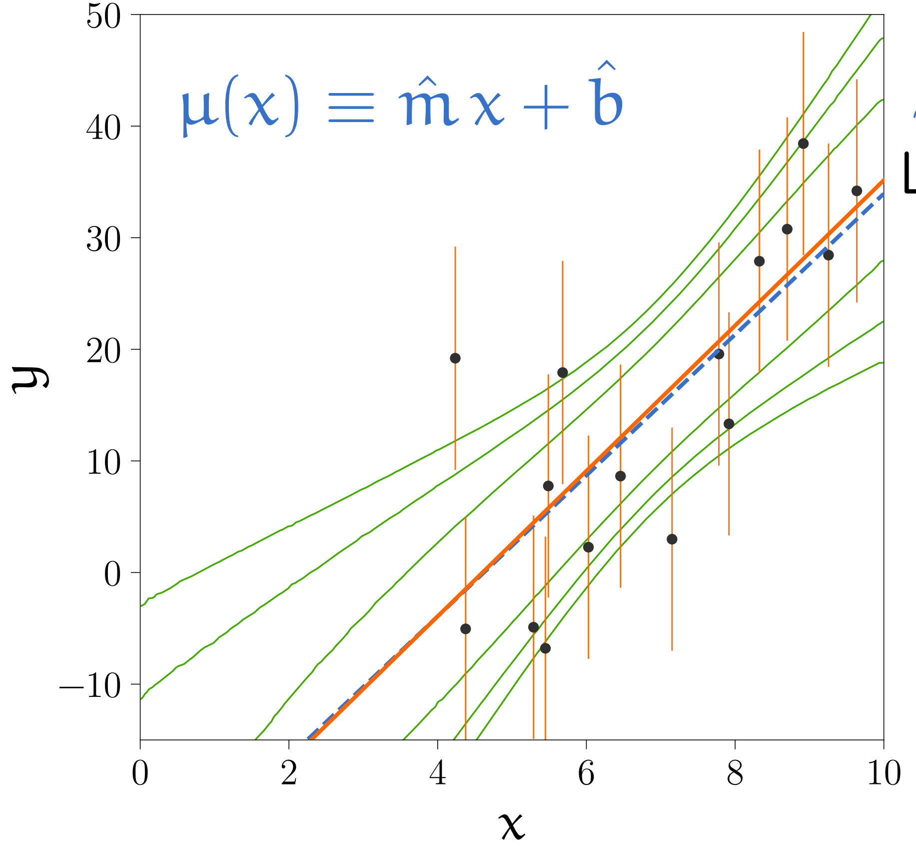
Verosimilitud marginalizada

$$\mathcal{L}(m \,|\, \{y\}, \{x\}, \{\sigma_y\}) = \int_b \mathcal{L}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})$$

$$\mathcal{L}(b \,|\, \{y\}, \{x\}, \{\sigma_y\}) = \int_m \mathcal{L}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})$$

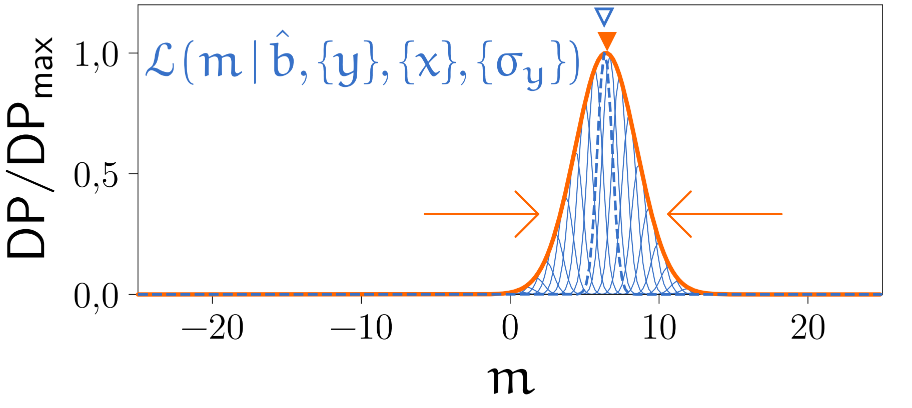


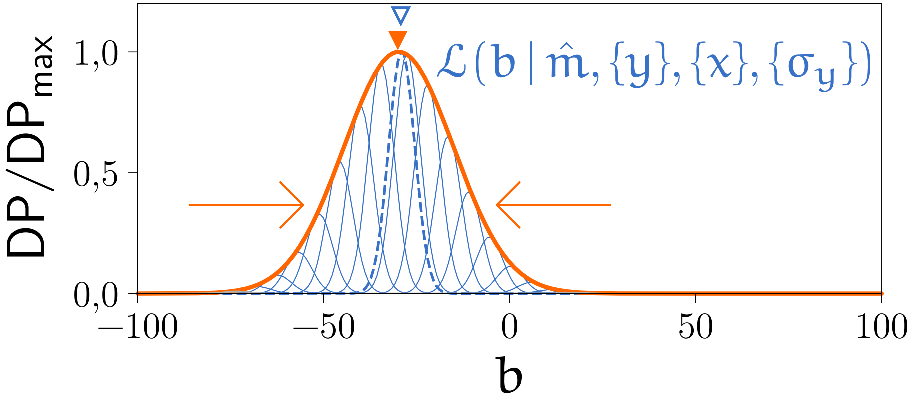
VEROSIMILITUD — REGRESIÓN LINEAL



$$\underbrace{\mathcal{L}_{\max} \equiv \mathcal{L}(\hat{m}, \hat{b} \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Máxima Verosimilitud}}$$

La marginalización es una forma de propagar incertidumbres





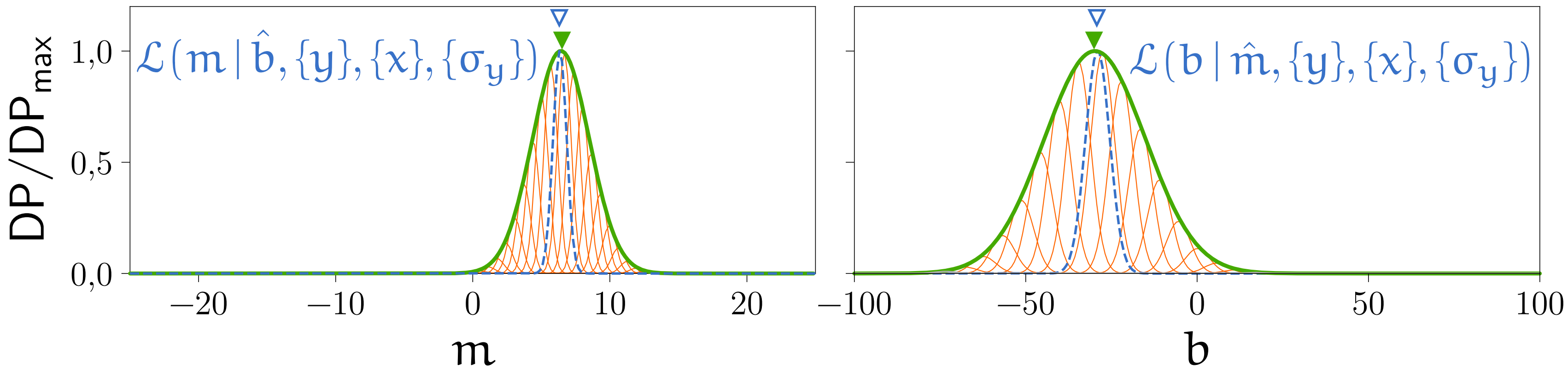
Verosimilitud marginalizada

$$\mathcal{L}(m \,|\, \{y\}, \{x\}, \{\sigma_y\}) = \int_b \overbrace{\mathcal{L}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})}^{\text{Verosimilitud}}$$

$$\mathcal{L}(b \,|\, \{y\}, \{x\}, \{\sigma_y\}) = \int_m \mathcal{L}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})$$

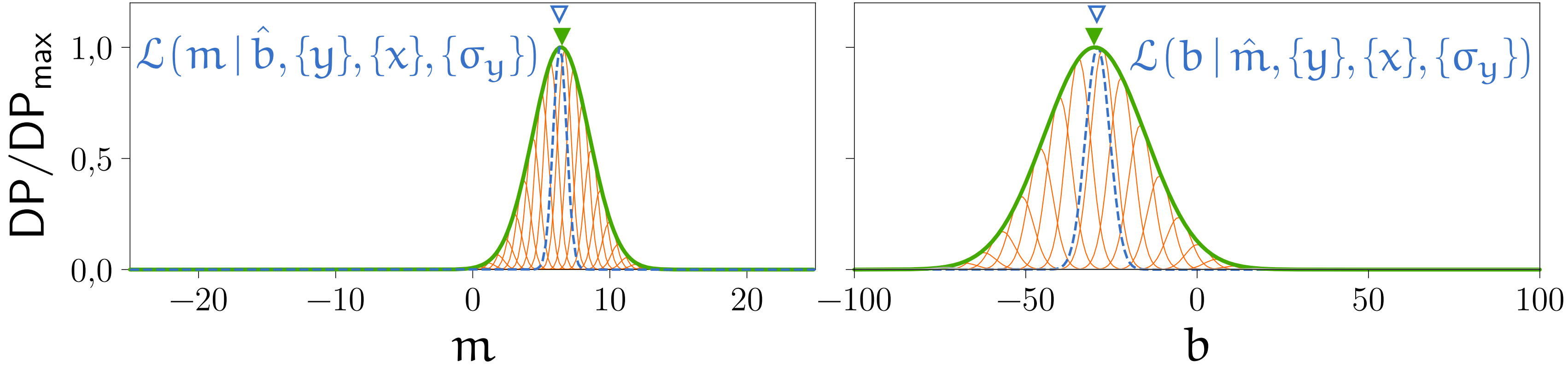
RESUMEN – VEROSIMILITUD

- El método de la Máxima Verosimilitud no informa sobre degeneraciones ni sobre incertidumbres.
- La Verosimilitud Condicional es buena en el límite de incertidumbres pequeñas en las observaciones.
- La marginalización es una forma de propagar las incertidumbres en el espacio de hipótesis.
- Un análisis completo (incertidumbres+degeneraciones) requiere del muestreo de la verosimilitud.



RESUMEN – VEROSIMILITUD

- El método de la Máxima Verosimilitud no informa sobre degeneraciones ni sobre incertidumbres.
- La Verosimilitud Condicional es buena en el límite de incertidumbres pequeñas en las observaciones.
- La marginalización es una forma de propagar las incertidumbres en el espacio de hipótesis.
- Un análisis completo (incertidumbres+degeneraciones) requiere del muestreo de la Verosimilitud.



¿Cuál es la mejor forma de muestrear la verosimilitud?

TEOREMA DE BAYES – INTERPRETACIÓN

La distribución previa se convierte en la distribución posterior

La observación actualiza el estado de la hipótesis

La verosimilitud pondera la distribución previa

$$\mathrm{P}(\mathcal{H} \mid \mathcal{D}, \mathcal{C}) = \frac{\mathcal{L}(\mathcal{H} \mid \mathcal{D}, \mathcal{C}) \times \mathrm{P}(\mathcal{H} \mid \mathcal{C})}{\mathrm{P}(\mathcal{D} \mid \mathcal{C})}$$

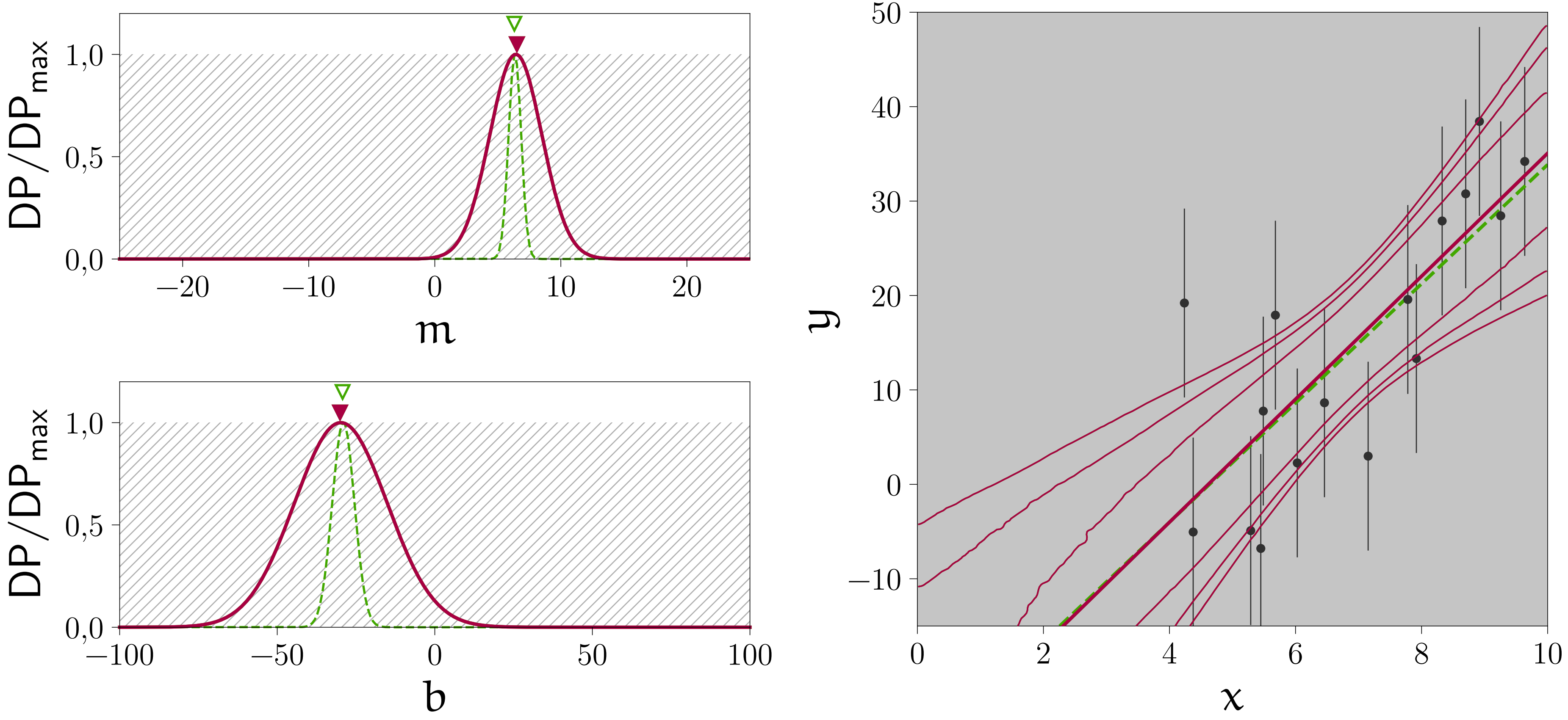
TEOREMA DE BAYES – OPERACIÓN COMPUTACIONAL

La muestra puede convertirse en muestra de la distribución posterior

Se muestrea la distribución previa

$$P(\mathcal{H} \mid \mathcal{D}, \mathcal{C}) = \frac{\mathcal{L}(\mathcal{H} \mid \mathcal{D}, \mathcal{C}) \times P(\mathcal{H} \mid \mathcal{C})}{P(\mathcal{D} \mid \mathcal{C})}$$

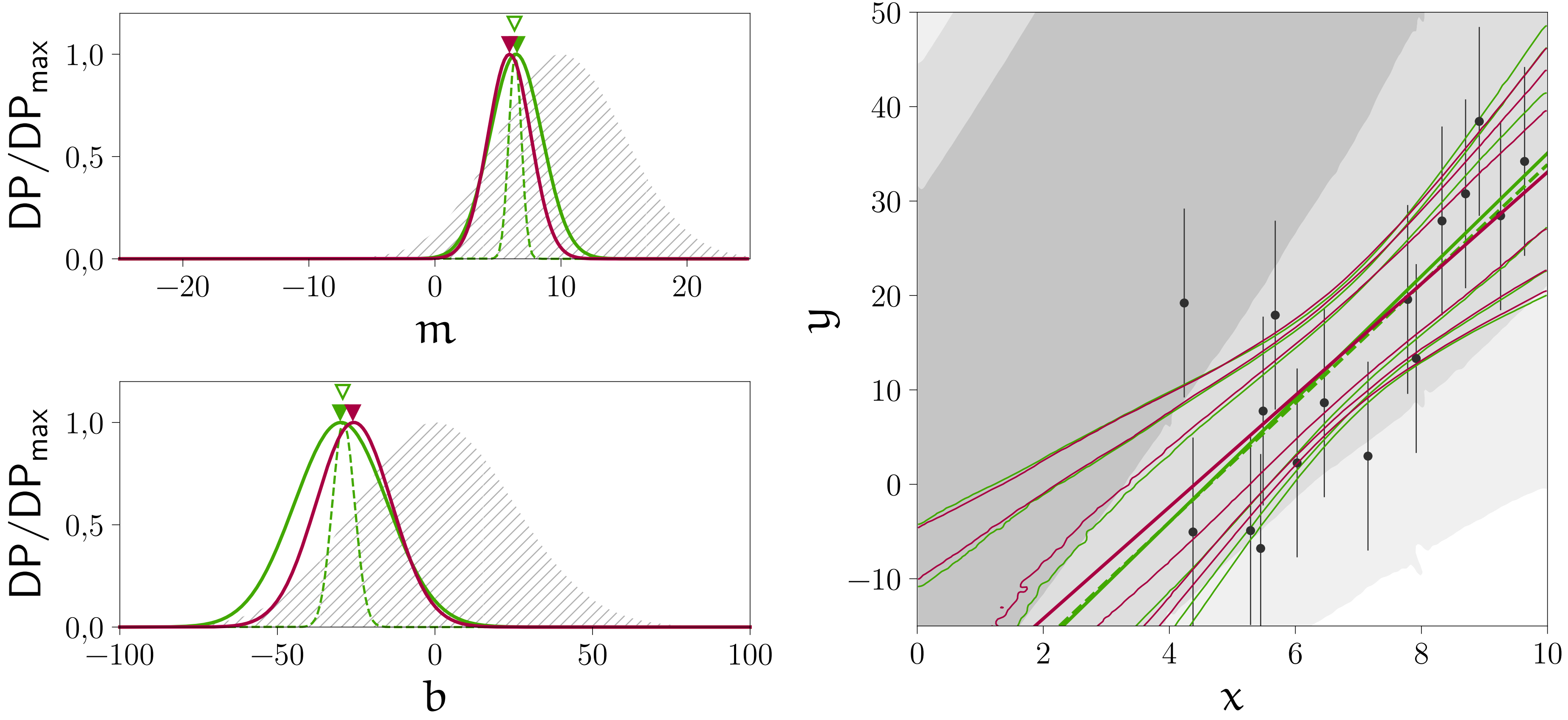
EL ROL DE LA DISTRIBUCIÓN PREVIA



Si la distribución previa es plana, la distribución posterior = verosimilitud.

$$\underbrace{P(m, b \mid \{y\}, \{x\}, \{\sigma_y\})}_{\text{Distribución Posterior}} \propto \underbrace{\mathcal{L}(m, b \mid \{y\}, \{x\}, \{\sigma_y\})}_{\text{Función de Verosimilitud}} \times \underbrace{P(m, b \mid \{x\}, \{\sigma_y\})}_{\text{Distribución Previa}}$$

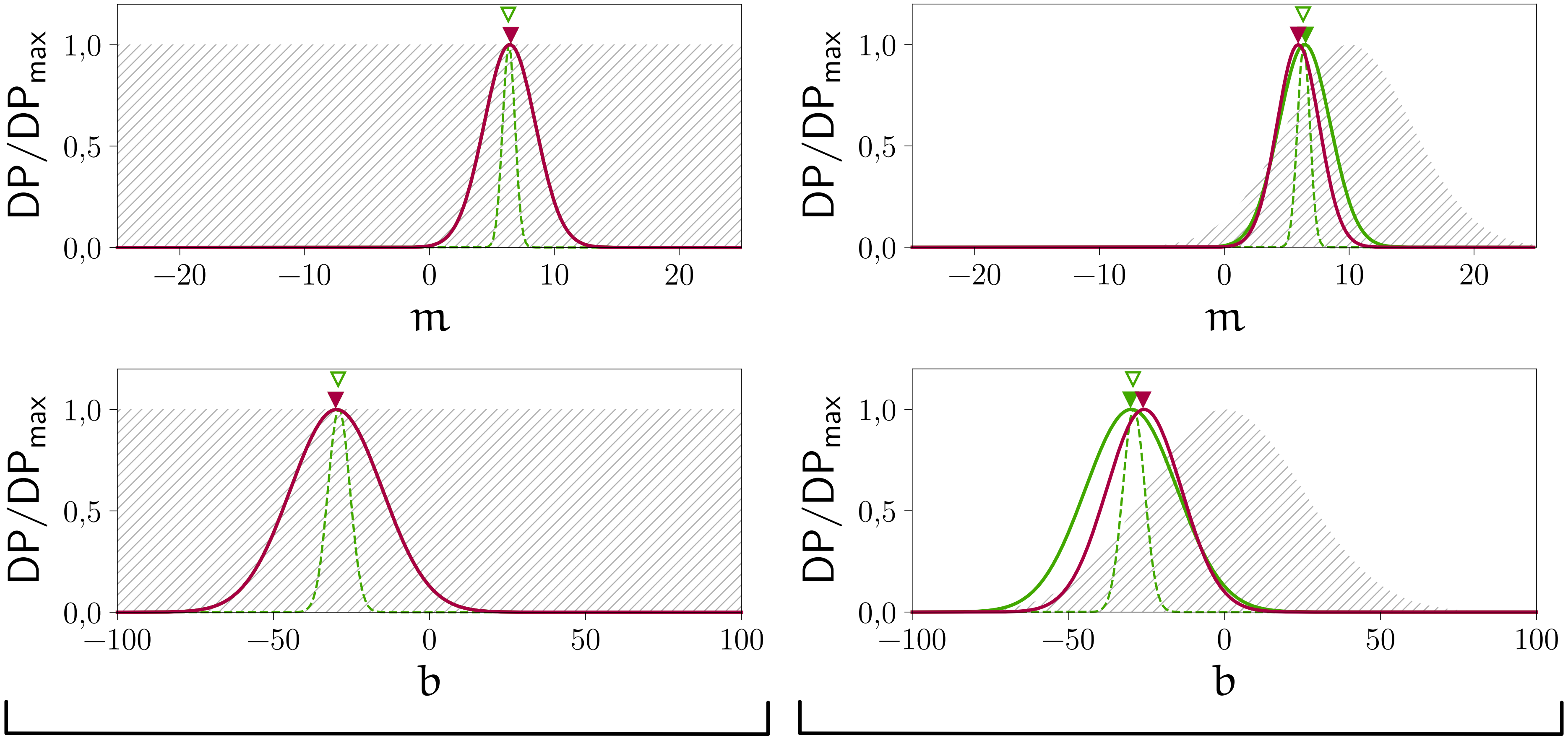
EL ROL DE LA DISTRIBUCIÓN PREVIA



¡Aún con una distribución previa informativa la distribución posterior puede ser similar a la verosimilitud!

$$\underbrace{P(m, b\,|\,\{y\},\{x\},\{\sigma_y\})}_{\text{Distribución Posterior}} \propto \underbrace{\mathcal{L}(m, b\,|\,\{y\},\{x\},\{\sigma_y\})}_{\text{Función de Verosimilitud}} \times \underbrace{P(m, b\,|\,\{x\},\{\sigma_y\})}_{\text{Distribución Previa}}$$

EL ROL DE LA DISTRIBUCIÓN PREVIA



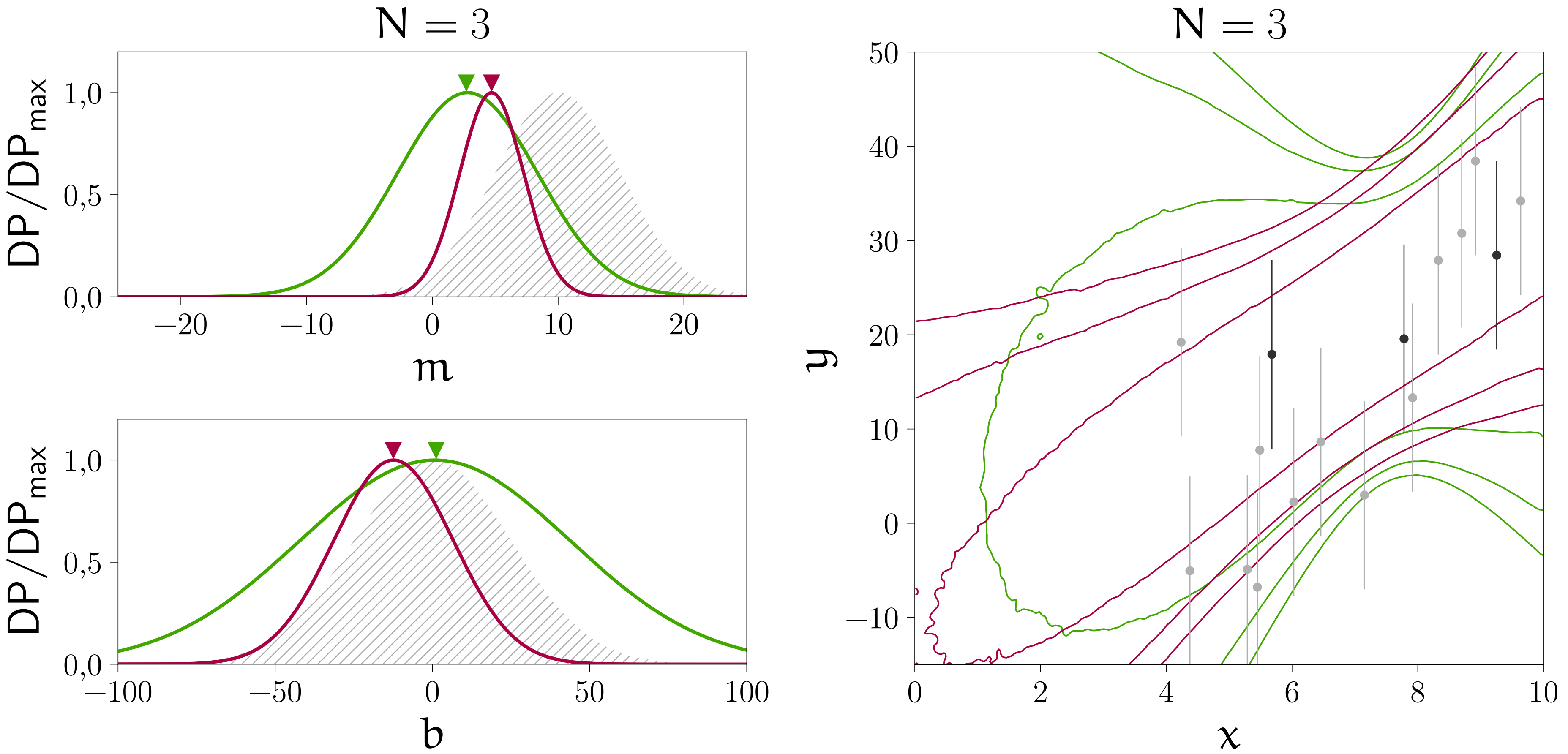
Distribución Previa Objetiva

Puede limitar el avance de la distribución posterior.

Distribución Previa Informativa

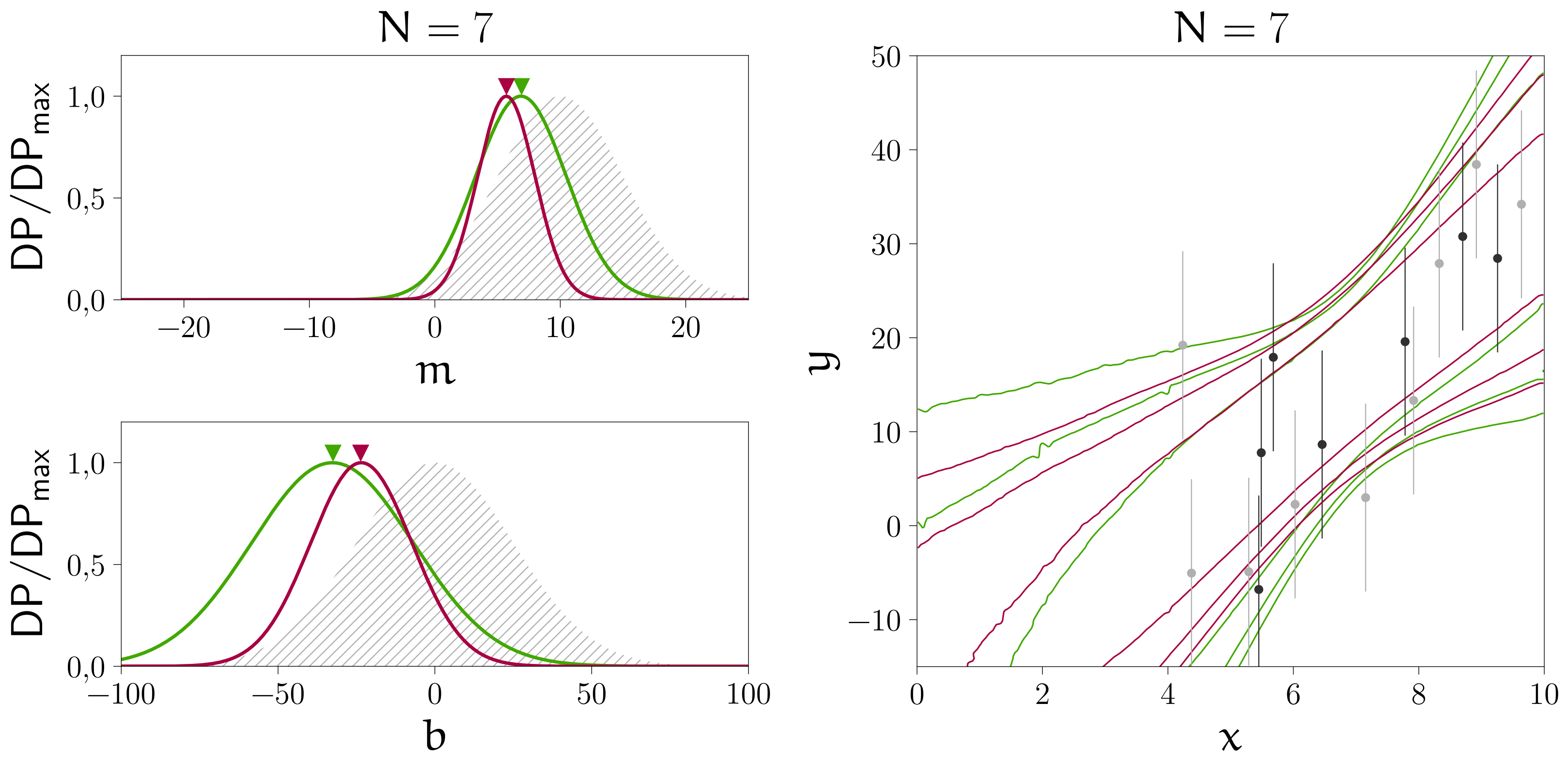
Puede plagar de prejuicios la distribución posterior.

EL ROL DE LA DISTRIBUCIÓN PREVIA — NÚMERO DE OBSERVACIONES



$$\underbrace{P(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Distribución Posterior}} \propto \underbrace{\mathcal{L}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Función de Verosimilitud}} \times \underbrace{P(m, b \,|\, \{x\}, \{\sigma_y\})}_{\text{Distribución Previa}}$$

EL ROL DE LA DISTRIBUCIÓN PREVIA – NÚMERO DE OBSERVACIONES



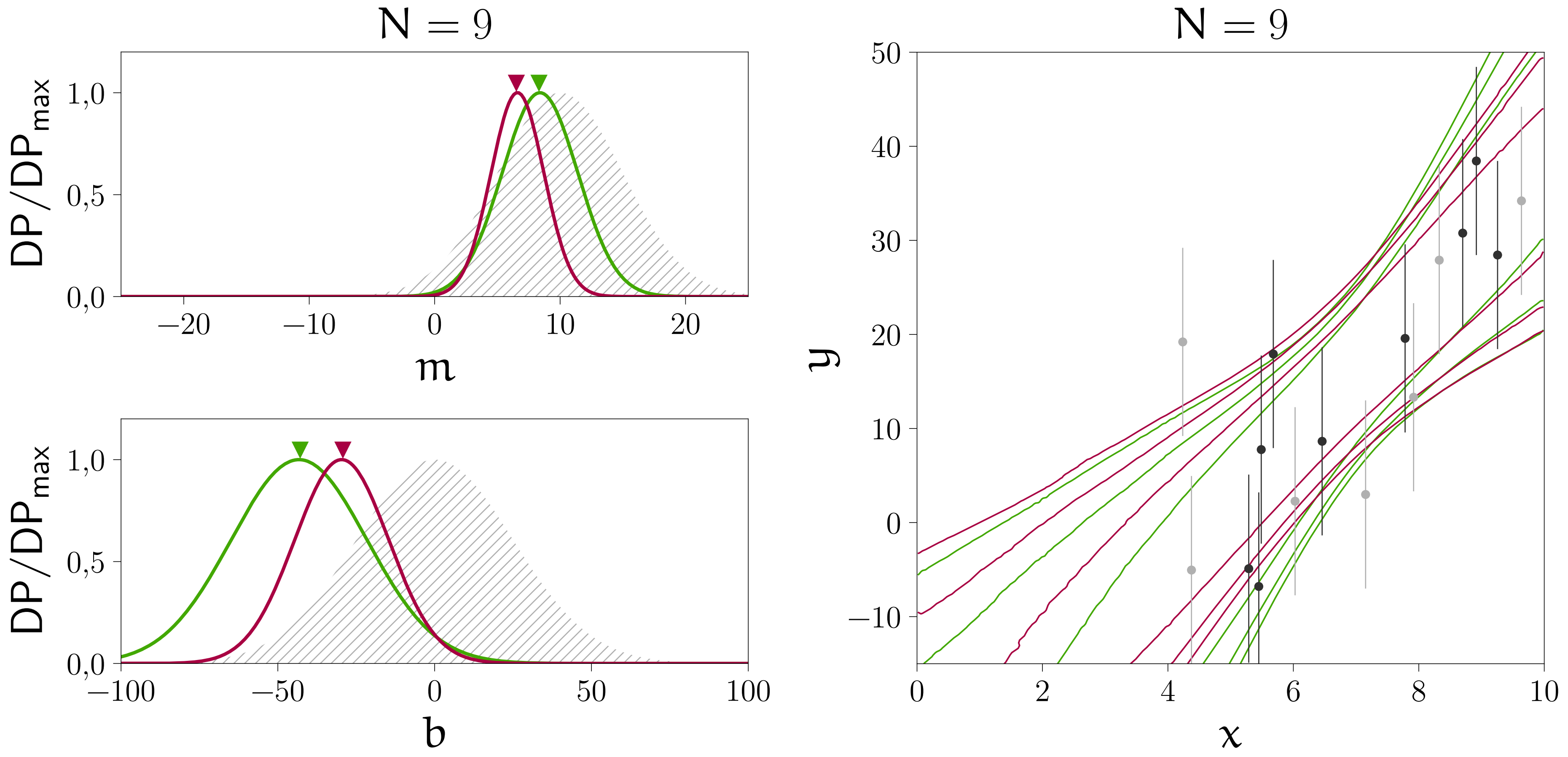
$$\mathrm{P}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\}) \propto \mathcal{L}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\}) \times \mathrm{P}(m, b \,|\, \{x\}, \{\sigma_y\})$$

Distribución Posterior

Función de Verosimilitud

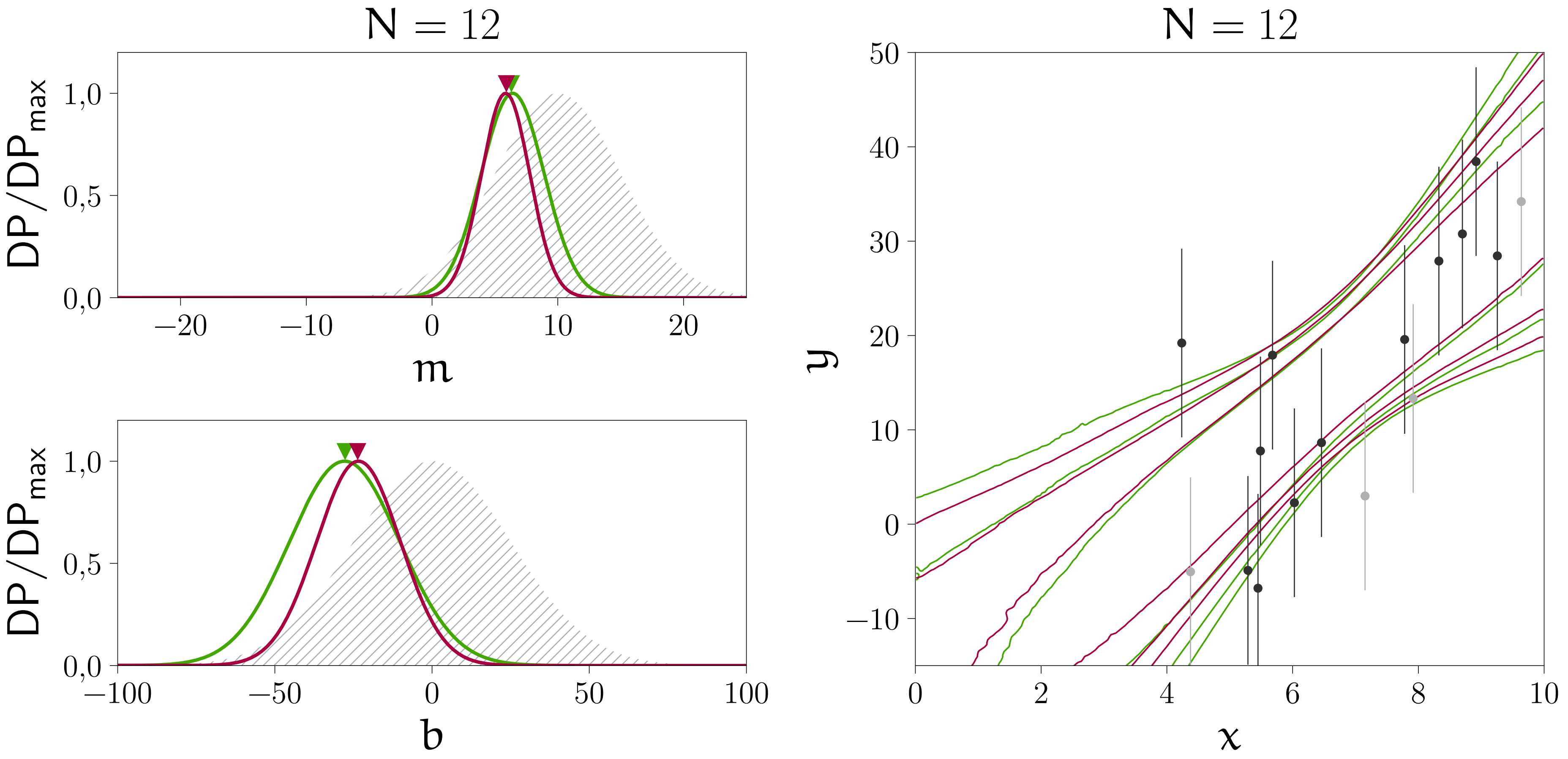
Distribución Previa

EL ROL DE LA DISTRIBUCIÓN PREVIA – NÚMERO DE OBSERVACIONES



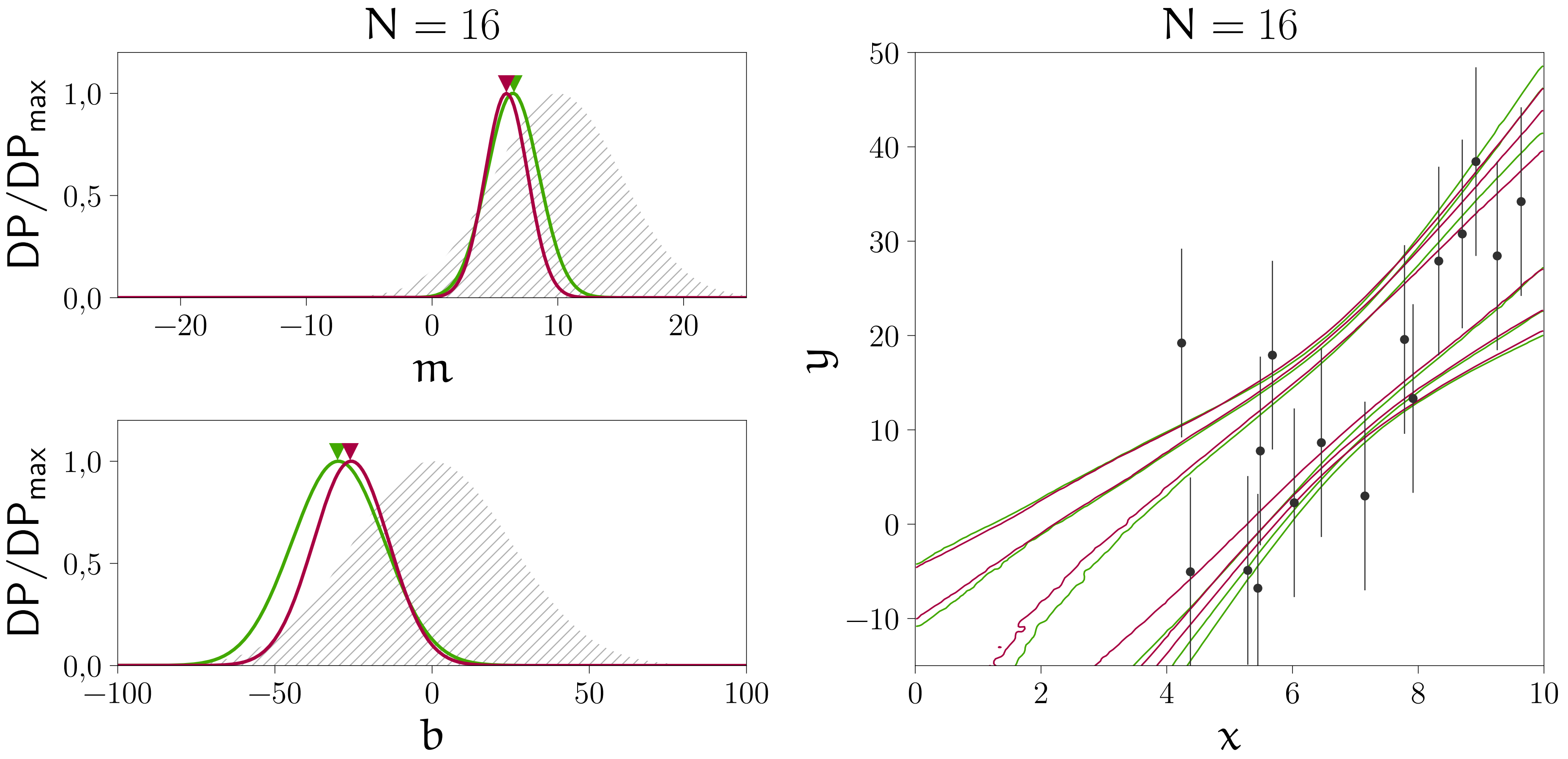
$$\underbrace{P(m, b \mid \{y\}, \{x\}, \{\sigma_y\})}_{\text{Distribución Posterior}} \propto \underbrace{\mathcal{L}(m, b \mid \{y\}, \{x\}, \{\sigma_y\})}_{\text{Función de Verosimilitud}} \times \underbrace{P(m, b \mid \{x\}, \{\sigma_y\})}_{\text{Distribución Previa}}$$

EL ROL DE LA DISTRIBUCIÓN PREVIA — NÚMERO DE OBSERVACIONES



$$\underbrace{P(m, b \mid \{y\}, \{x\}, \{\sigma_y\})}_{\text{Distribución Posterior}} \propto \underbrace{\mathcal{L}(m, b \mid \{y\}, \{x\}, \{\sigma_y\})}_{\text{Función de Verosimilitud}} \times \underbrace{P(m, b \mid \{x\}, \{\sigma_y\})}_{\text{Distribución Previa}}$$

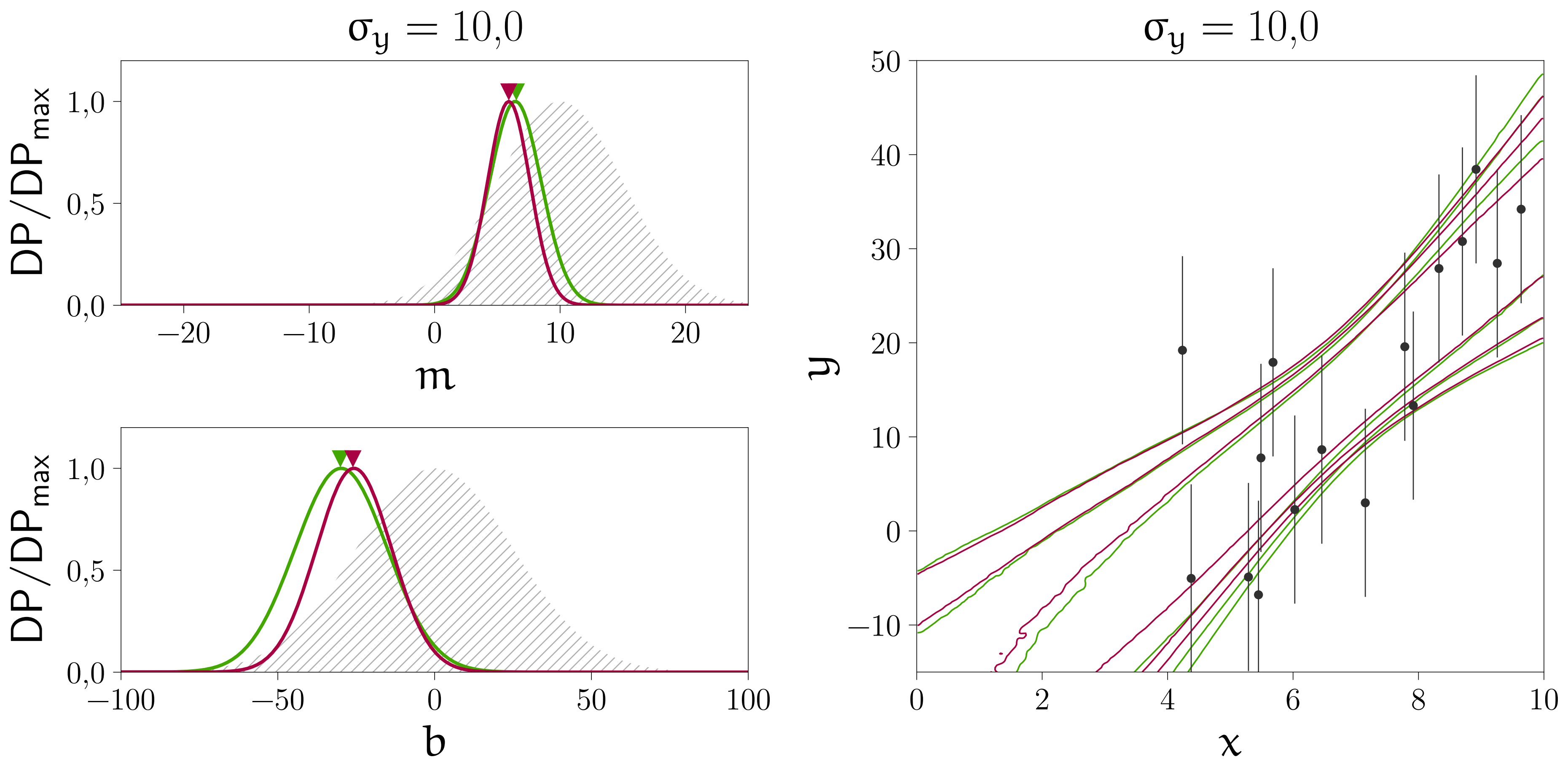
EL ROL DE LA DISTRIBUCIÓN PREVIA — NÚMERO DE OBSERVACIONES



En el límite N⟶∞ la distribución previa se vuelve irrelevante.

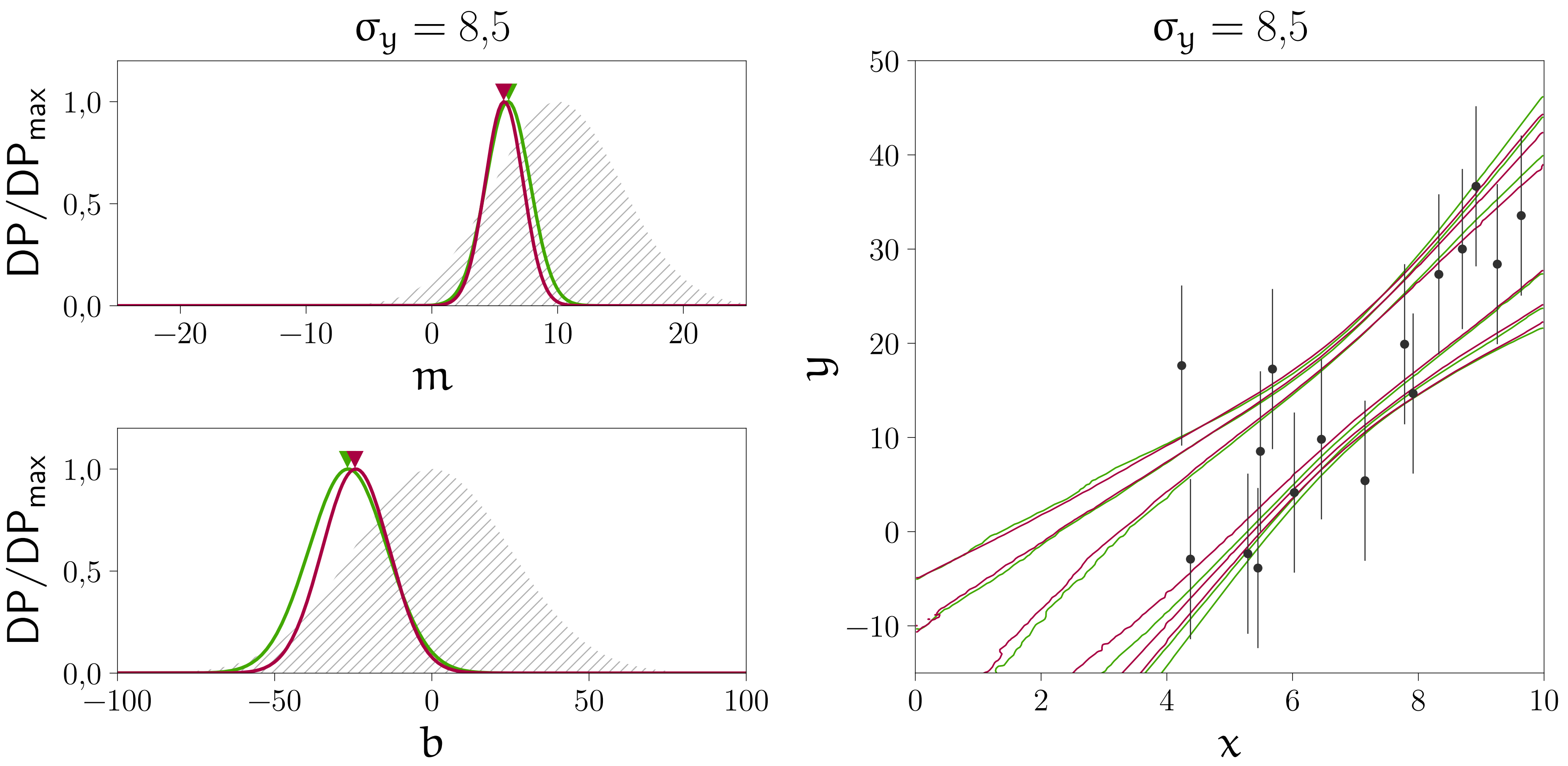
$$\underbrace{P(m, b \mid \{y\}, \{x\}, \{\sigma_y\})}_{\text{Distribución Posterior}} \propto \underbrace{\mathcal{L}(m, b \mid \{y\}, \{x\}, \{\sigma_y\})}_{\text{Función de Verosimilitud}} \times \underbrace{P(m, b \mid \{x\}, \{\sigma_y\})}_{\text{Distribución Previa}}$$

EL ROL DE LA DISTRIBUCIÓN PREVIA — SEÑAL/RUIDO DE LAS OBSERVACIONES



$$\underbrace{P(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Distribución Posterior}} \propto \underbrace{\mathcal{L}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Función de Verosimilitud}} \times \underbrace{P(m, b \,|\, \{x\}, \{\sigma_y\})}_{\text{Distribución Previa}}$$

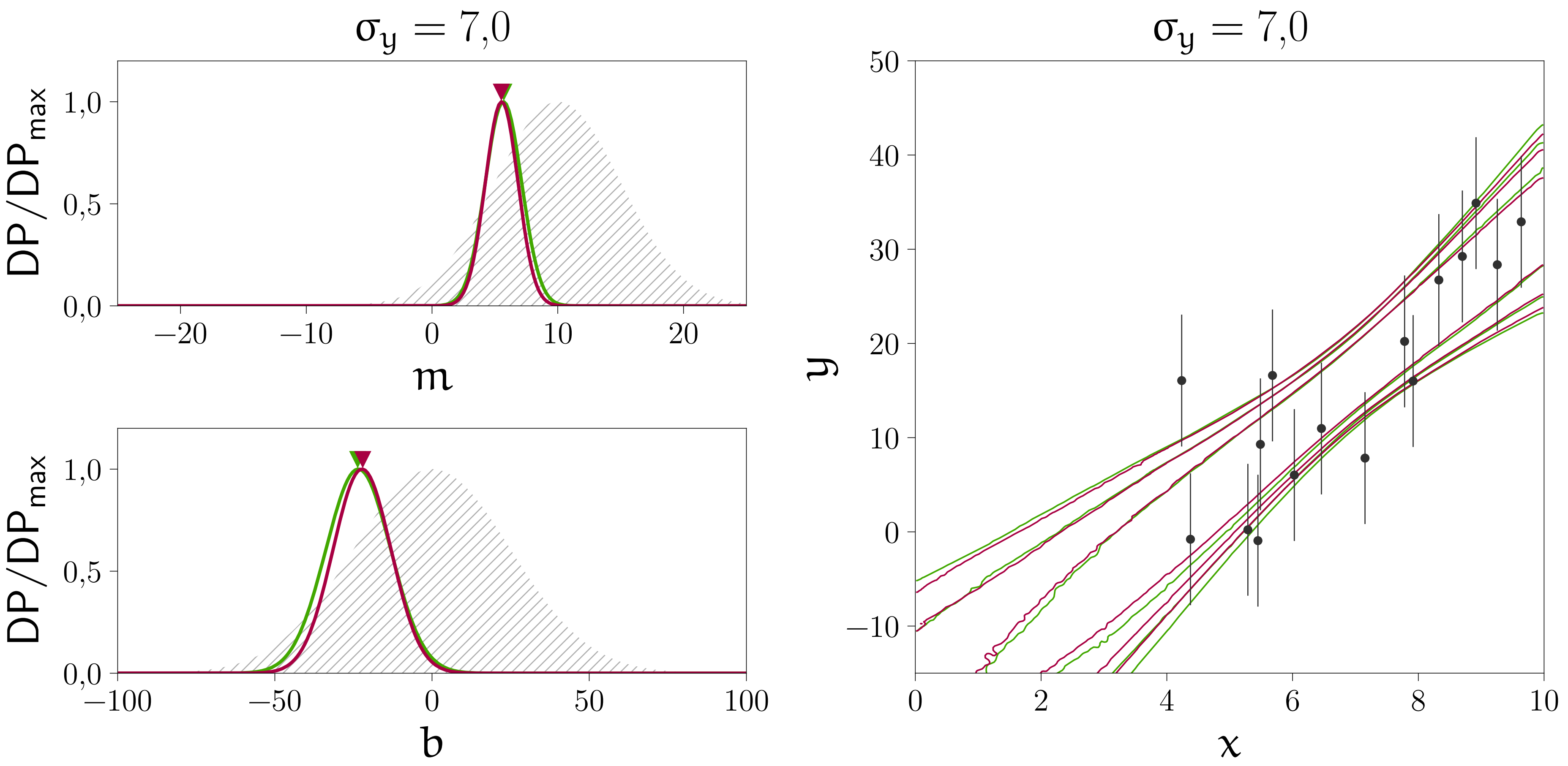
EL ROL DE LA DISTRIBUCIÓN PREVIA — SEÑAL/RUIDO DE LAS OBSERVACIONES



$$P(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\}) \propto \mathcal{L}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\}) \times P(m, b \,|\, \{x\}, \{\sigma_y\})$$

Distribución Posterior — Función de Verosimilitud — Distribución Previa

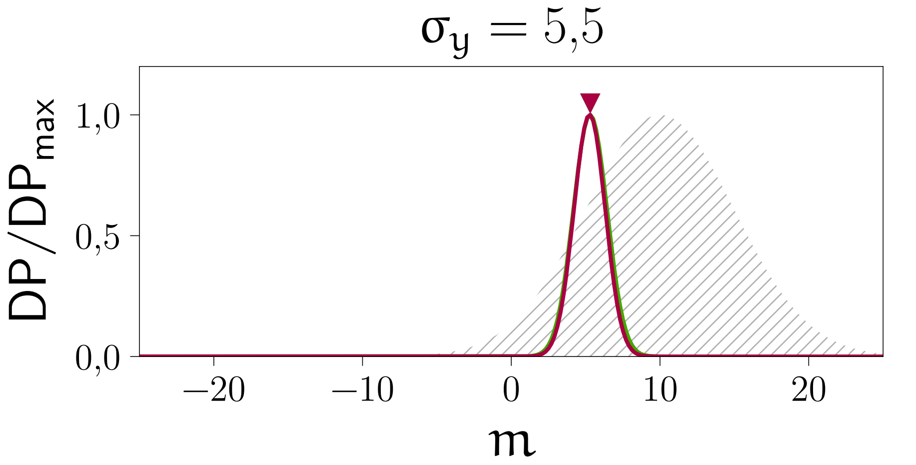
EL ROL DE LA DISTRIBUCIÓN PREVIA — SEÑAL/RUIDO DE LAS OBSERVACIONES

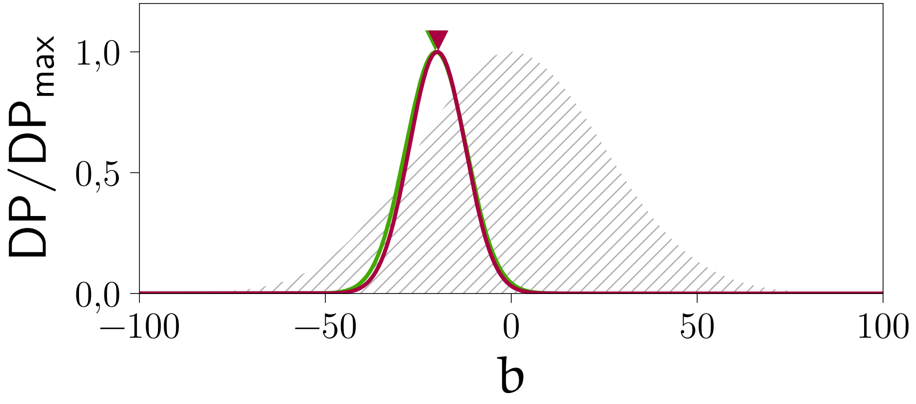


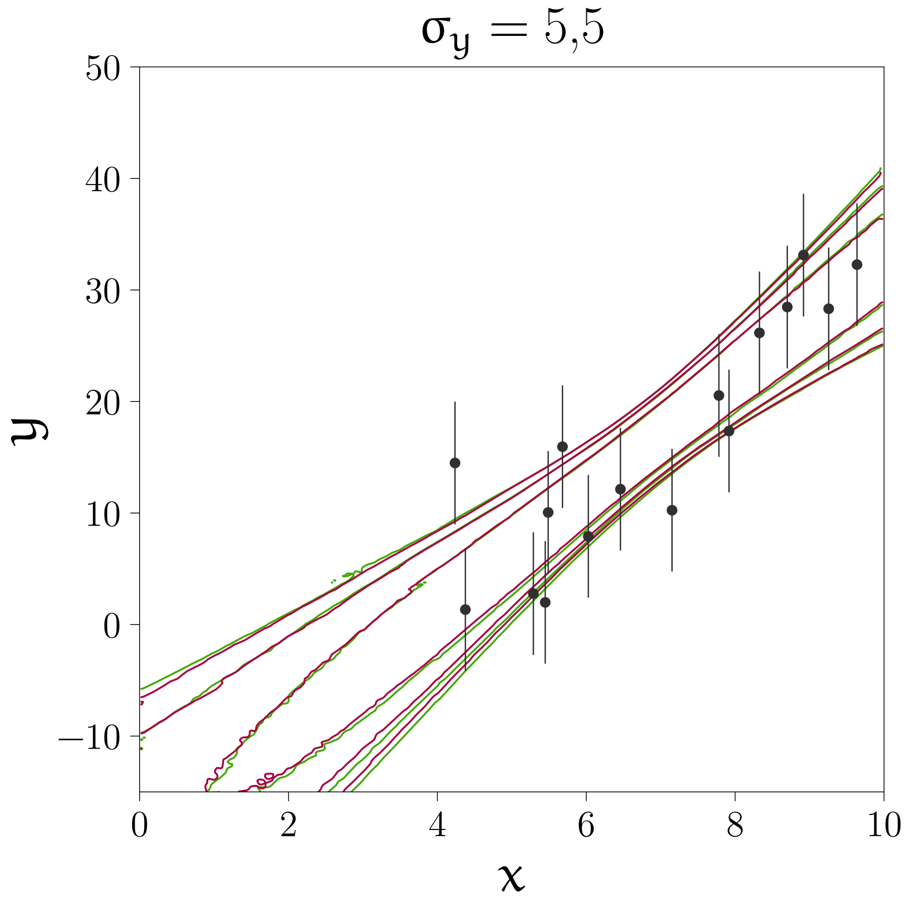
$$P(m, b \mid \{y\}, \{x\}, \{\sigma_y\}) \propto \mathcal{L}(m, b \mid \{y\}, \{x\}, \{\sigma_y\}) \times P(m, b \mid \{x\}, \{\sigma_y\})$$

Distribución Posterior — Función de Verosimilitud — Distribución Previa

EL ROL DE LA DISTRIBUCIÓN PREVIA — SEÑAL/RUIDO DE LAS OBSERVACIONES

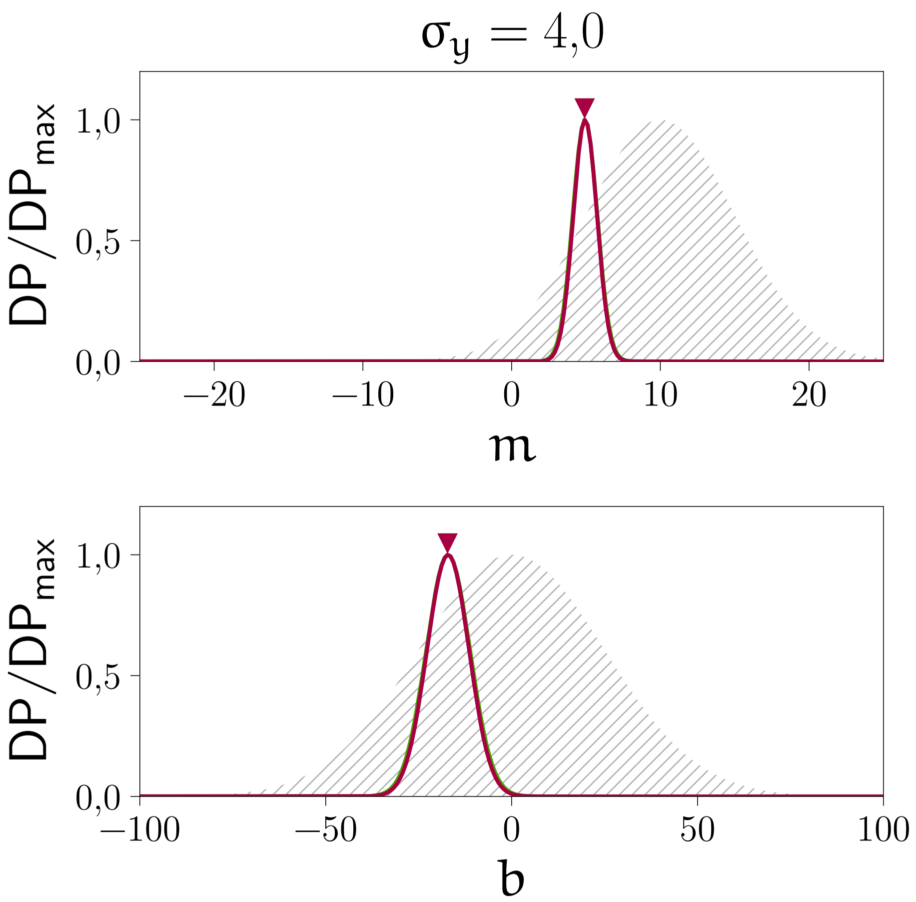


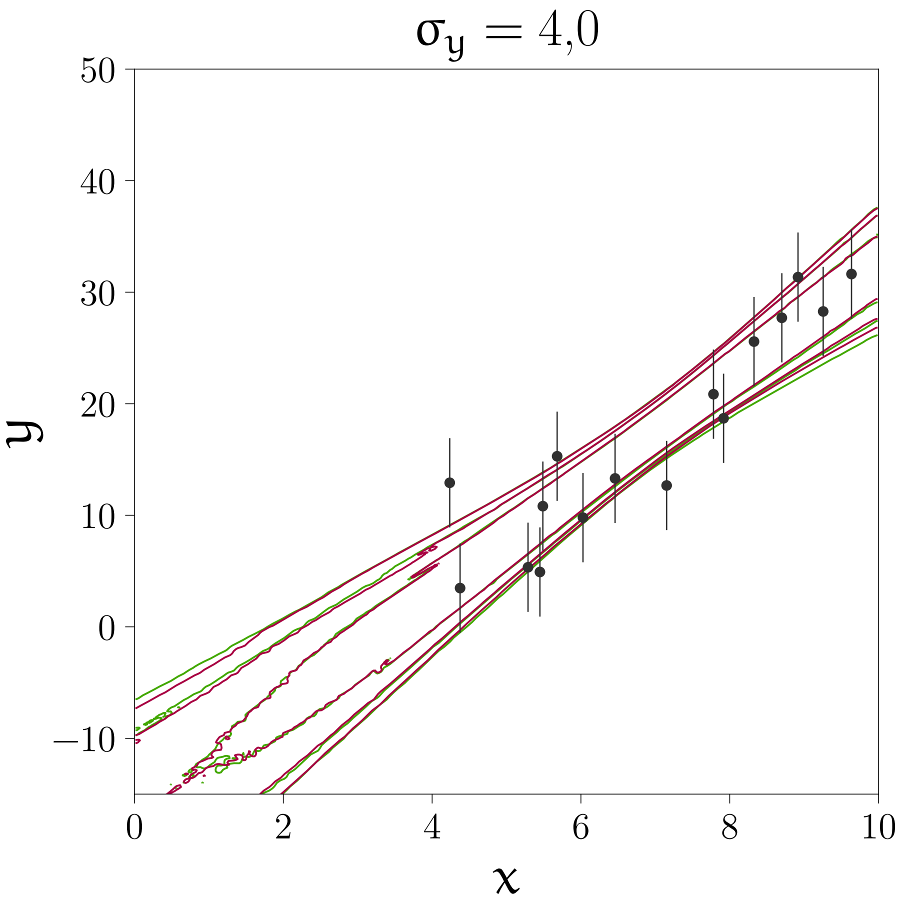




$$\underbrace{P(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Distribución Posterior}} \propto \underbrace{\mathcal{L}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Función de Verosimilitud}} \times \underbrace{P(m, b \,|\, \{x\}, \{\sigma_y\})}_{\text{Distribución Previa}}$$

EL ROL DE LA DISTRIBUCIÓN PREVIA — SEÑAL/RUIDO DE LAS OBSERVACIONES



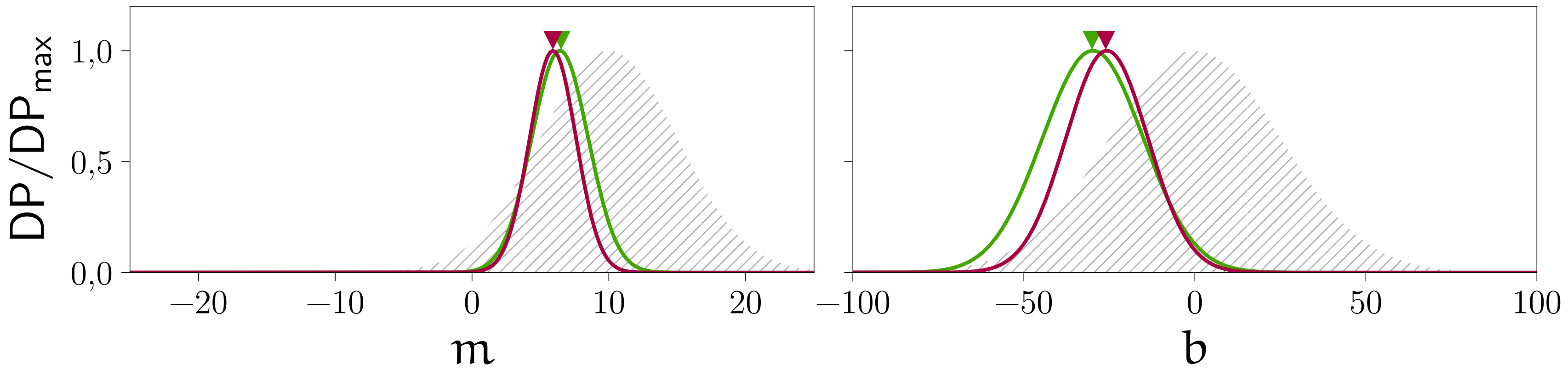


En el límite $\sigma_y \longrightarrow 0$ la distribución previa se vuelve irrelevante.

$$\underbrace{P(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Distribución Posterior}} \propto \underbrace{\mathcal{L}(m, b \,|\, \{y\}, \{x\}, \{\sigma_y\})}_{\text{Función de Verosimilitud}} \times \underbrace{P(m, b \,|\, \{x\}, \{\sigma_y\})}_{\text{Distribución Previa}}$$

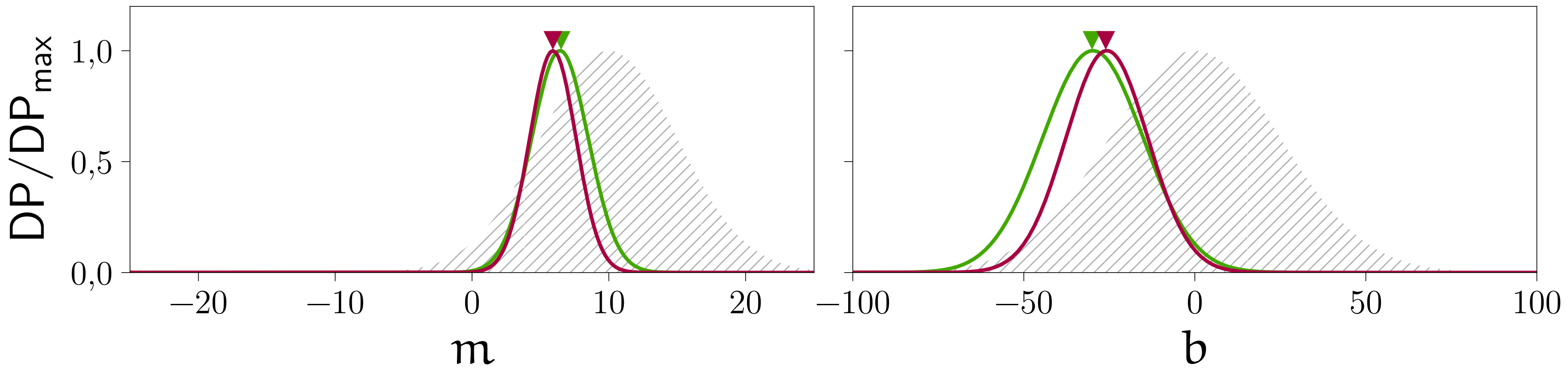
RESUMEN — EL ROL DE LA DISTRIBUCIÓN PREVIA

- Proporcionar un punto de partida en el espacio de Hipótesis.
- Facilitar el muestreo de la Verosimilitud.
- En el límite de observaciones informativas, la Distribución Previa debe ser irrelevante.
- Aún así, la Distribución Previa debe ser independiente de las observaciones.



RESUMEN – EL ROL DE LA DISTRIBUCIÓN PREVIA

- Proporcionar un punto de partida en el espacio de Hipótesis.
- Facilitar el muestreo de la Verosimilitud.
- En el límite de observaciones informativas, la Distribución Previa debe ser irrelevante.
- Aún así, la Distribución Previa debe ser independiente de las observaciones.



¿Cómo asignar la Distribución Previa de manera honesta?

¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?

$$\mathcal{L}(\mathcal{H} \mid \mathcal{D}, \mathcal{C}),\ \mathrm{P}(\mathcal{H} \mid \mathcal{C}) \xrightarrow{?} \text{Información}$$

¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?

$$\mathcal{L}(\mathcal{H} \mid \mathcal{D}, \mathcal{C}),\ P(\mathcal{H} \mid \mathcal{C}) \xrightarrow{?} \text{Información}$$

$$P(N \mid \mathcal{C})$$

Probabilidad de nubes
sobre Sta. Rosa

¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?

$$\mathcal{L}(\mathcal{H} \mid \mathcal{D}, \mathcal{C}),\ P(\mathcal{H} \mid \mathcal{C}) \xrightarrow{?} \text{Información}$$

$$P(N \mid \mathcal{C})$$

Probabilidad de nubes
sobre Sta. Rosa

$$I(P) \to 0$$

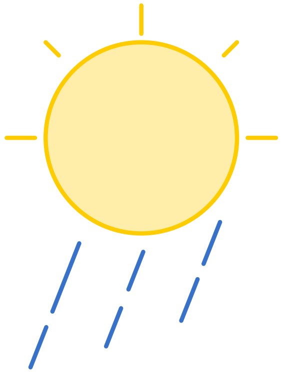
¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?

$$\mathcal{L}(\mathcal{H} \mid \mathcal{D}, \mathcal{C}),\ P(\mathcal{H} \mid \mathcal{C}) \xrightarrow{?} \text{Información}$$

$$P(N \mid \mathcal{C})$$

Probabilidad de nubes
sobre Sta. Rosa

$$I(P) \to 0$$

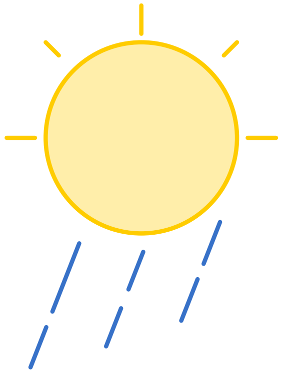


$$P(L, \bar{N} \mid \mathcal{C})$$

Probabilidad de lluvia
con cielo despejado

¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?

$$\mathcal{L}(\mathcal{H} \mid \mathcal{D}, \mathcal{C}),\ P(\mathcal{H} \mid \mathcal{C}) \xrightarrow{?} \text{Información}$$



$$P(N \mid \mathcal{C})$$

Probabilidad de nubes
sobre Sta. Rosa

$$I(P) \to 0$$

$$P(L, \bar{N} \mid \mathcal{C})$$

Probabilidad de lluvia
con cielo despejado

$$I(P) \to \infty$$

¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?

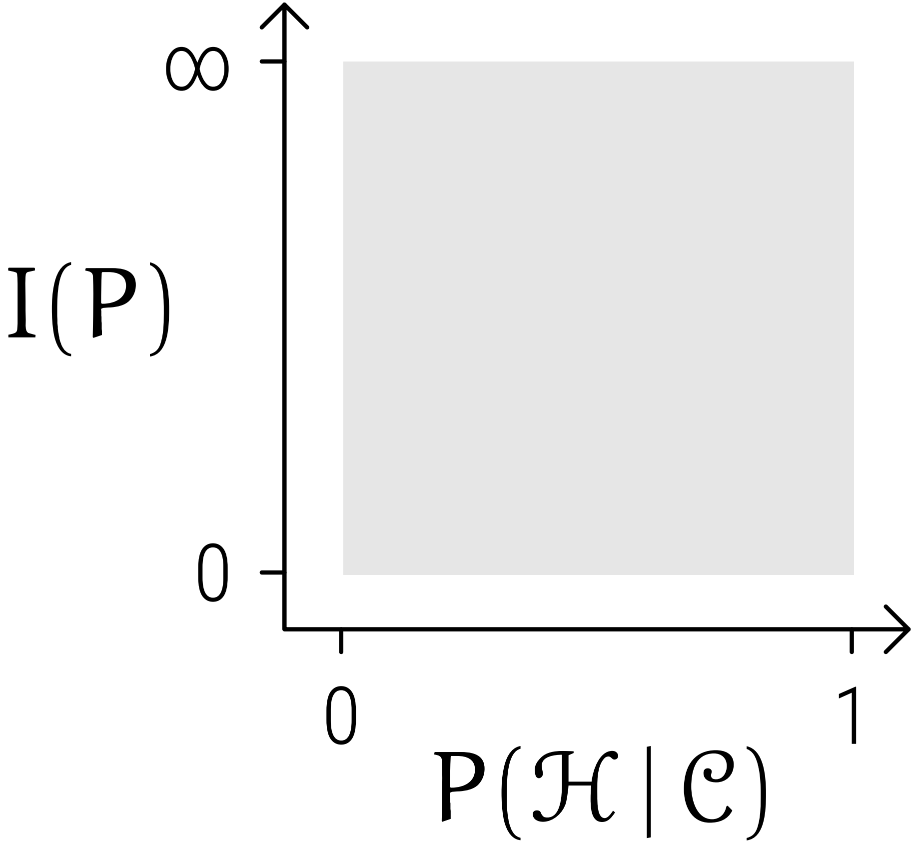
INFORMACIÓN: medida de la impredictibilidad en la ocurrencia de un evento.

¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?

INFORMACIÓN: medida de la impredictibilidad en la ocurrencia de un evento.

REGLAS DE SHANNON:

$$\mathrm{I}: \ \mathrm{I}(\mathrm{P})$$



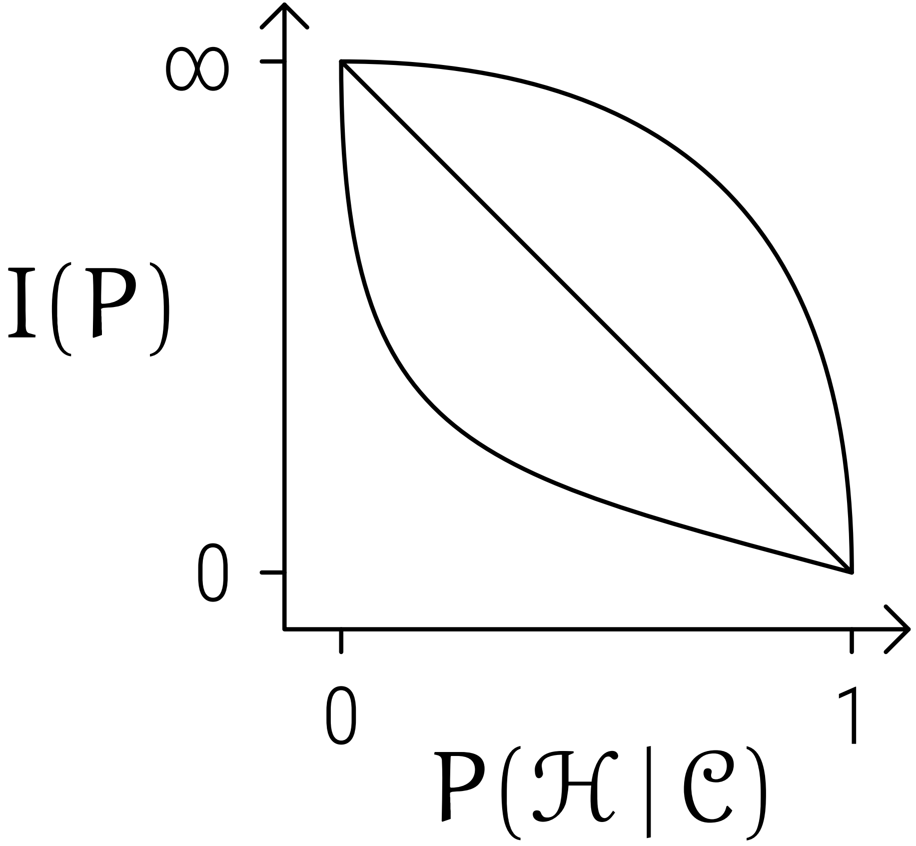
¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?

INFORMACIÓN: medida de la impredictibilidad en la ocurrencia de un evento.

REGLAS DE SHANNON:

$\mathrm{I}: \ \mathrm{I}(\mathrm{P})$

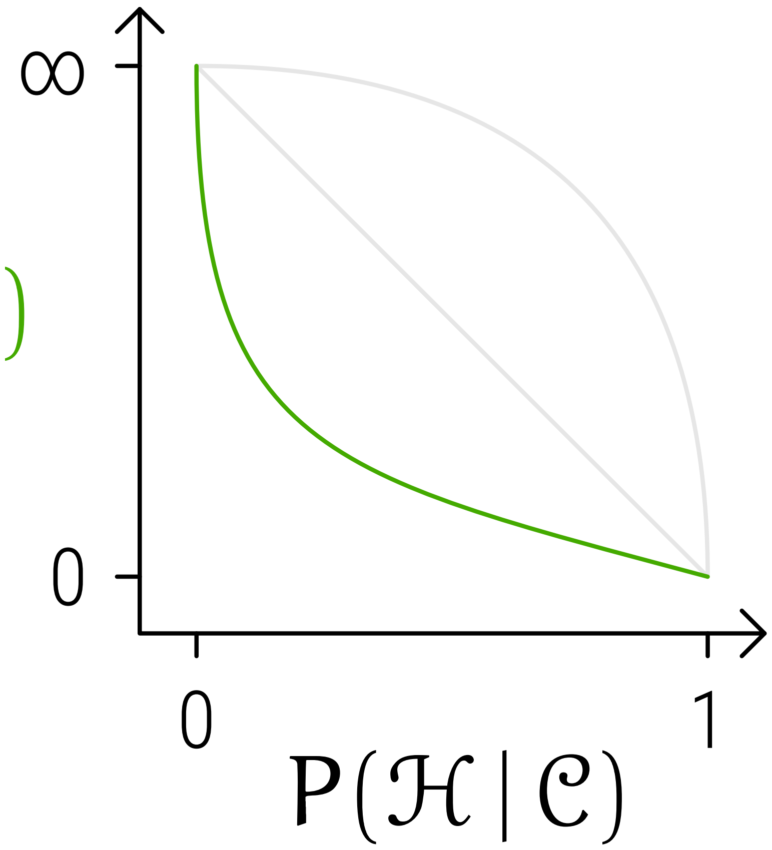
$\mathrm{I}(\mathrm{P})$ es monótona y continua



¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?

INFORMACIÓN: medida de la impredictibilidad en la ocurrencia de un evento.

$$-\log P \equiv I(P)$$



REGLAS DE SHANNON:

$$I : I(P)$$

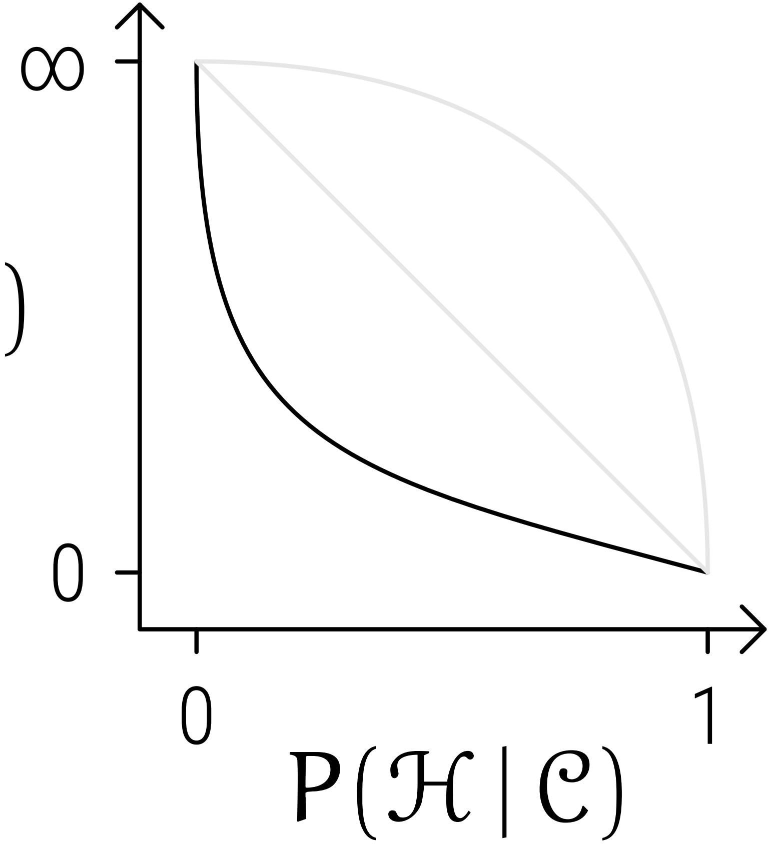
$I(P)$ es monótona y continua

$$I(P_A \times P_B) = I(P_A) + I(P_B)$$

¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?

INFORMACIÓN: medida de la impredictibilidad en la ocurrencia de un evento.

$$-\log \mathrm{P} \equiv \mathrm{I}(\mathrm{P})$$



REGLAS DE SHANNON:

$$\mathrm{I} : \mathrm{I}(\mathrm{P})$$

$\mathrm{I}(\mathrm{P})$ es monótona y continua

$$\mathrm{I}(\mathrm{P}_A \times \mathrm{P}_B) = \mathrm{I}(\mathrm{P}_A) + \mathrm{I}(\mathrm{P}_B)$$

$\underbrace{\mathrm{P}(\mathcal{H} \,|\, \mathcal{C})}_{\text{Distribución Previa}}$

- Muda respecto de $\mathcal{L}(\mathcal{H} \,|\, \mathcal{D}, \mathcal{C})$
- Informativa únicamente a través de $\mathcal{C}$

¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?: MÉTODO DE MÁXIMA ENTROPÍA

$$\underbrace{L(P, P'; \mathcal{H})}_{\text{Información esperada ó Entropía, H(P)}} \equiv \int_{\mathcal{H}} \underbrace{P(\mathcal{H} \,|\, \mathcal{C})}_{\text{Distribución de Probabilidad a priori}} \times \underbrace{\log 1/P(\mathcal{H} \,|\, \mathcal{C})}_{\text{Información adquirida tras la observación}}$$

¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?: MÉTODO DE MÁXIMA ENTROPÍA

$$\underbrace{L(P, P'; \mathcal{H})}_{\text{Información esperada ó Entropía, H(P)}} \equiv \int_{\mathcal{H}} \underbrace{P(\mathcal{H} \,|\, \mathcal{C})}_{\text{Distribución de Probabilidad a priori}} \times \underbrace{\log 1/P(\mathcal{H} \,|\, \mathcal{C})}_{\text{Información adquirida tras la observación}} \; ; \; \underbrace{\sum_{i=0} \lambda_i \, g_i(P, P'; \mathcal{H})}_{\text{Restricciones impuestas por la información contrastable}}$$

Información esperada ó Entropía, H(P)

Distribución de Probabilidad *a priori*

Información adquirida tras la observación

Restricciones impuestas por la información contrastable

¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?: MÉTODO DE MÁXIMA ENTROPÍA

La información verificable define el conjunto de restricciones

Las restricciones fijan propiedades macro

$$\underbrace{L(P, P'; \mathcal{H})}_{\text{Información esperada ó Entropía, H(P)}} \equiv \int_{\mathcal{H}} \underbrace{P(\mathcal{H} \mid \mathcal{C})}_{\text{\textit{Distribución de} Probabilidad \textit{a priori}}} \times \underbrace{\log 1/P(\mathcal{H} \mid \mathcal{C})}_{\text{Información adquirida tras la observación}} \; ; \; \underbrace{\sum_{i=0} \lambda_i \, g_i(P, P'; \mathcal{H})}_{\text{Restricciones impuestas por la información contrastable}}$$

¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?: MÉTODO DE MÁXIMA ENTROPÍA

La información verificable define el conjunto de restricciones

Las restricciones fijan propiedades macro

$$\underbrace{L(P, P'; \mathcal{H})}_{\text{Información esperada ó Entropía, H(P)}} \equiv \int_{\mathcal{H}} \underbrace{P(\mathcal{H} \,|\, \mathcal{C})}_{\text{Distribución de Probabilidad a priori}} \times \underbrace{\log 1/P(\mathcal{H} \,|\, \mathcal{C})}_{\text{Información adquirida tras la observación}} \; ; \; \underbrace{\sum_{i=0} \lambda_i \, g_i(P, P'; \mathcal{H})}_{\text{Restricciones impuestas por la información contrastable}}$$

$$\max H : \underbrace{\frac{\partial L}{\partial P} - \frac{d}{d\mathcal{H}} \frac{\partial L}{\partial P'} + \sum_i \lambda_i \frac{\partial g_i}{\partial P} = 0}_{\text{Ecuación de Euler-Lagrange}}$$

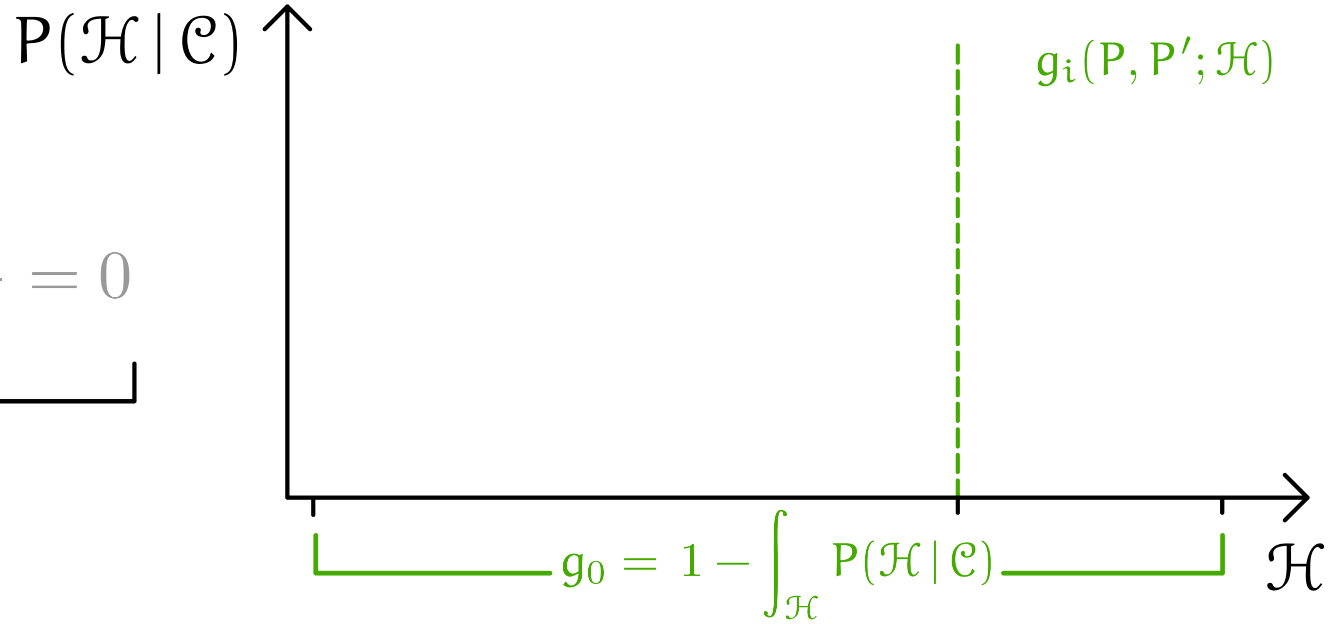
¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?: MÉTODO DE MÁXIMA ENTROPÍA

Las restricciones fijan propiedades macro

La información verificable define el conjunto de restricciones

$$\underbrace{L(P,P';\mathcal{H})}_{\text{Información esperada ó Entropía, H(P)}} \equiv \int_{\mathcal{H}} \underbrace{P(\mathcal{H}\,|\,\mathcal{C})}_{\text{Distribución de Probabilidad a priori}} \times \underbrace{\log 1/P(\mathcal{H}\,|\,\mathcal{C})}_{\text{Información adquirida tras la observación}}\;;\; \underbrace{\sum_{i=0} \lambda_i\, g_i(P,P';\mathcal{H})}_{\text{Restricciones impuestas por la información contrastable}}$$

$$\max \mathrm{H} : \underbrace{\frac{\partial L}{\partial P} - \frac{d}{d\mathcal{H}}\frac{\partial L}{\partial P'} + \sum_i \lambda_i \frac{\partial g_i}{\partial P} = 0}_{\text{Ecuación de Euler-Lagrange}}$$



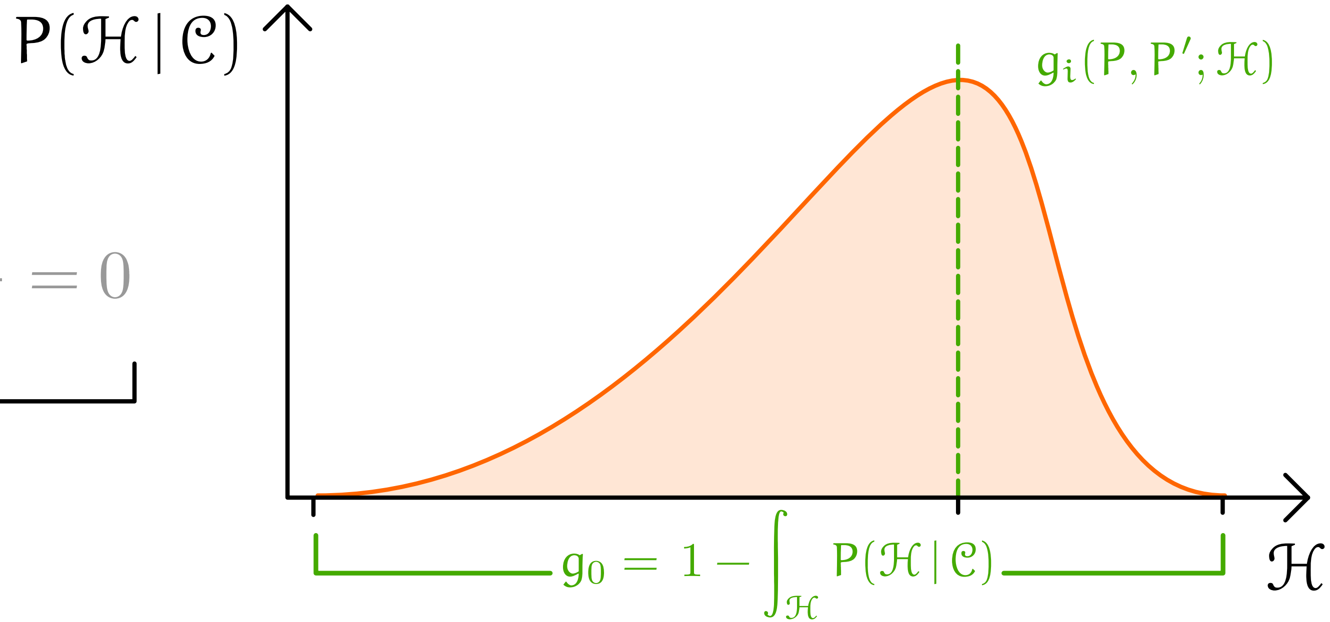
¿CÓMO ASIGNAR LA DISTRIBUCIÓN PREVIA?: MÉTODO DE MÁXIMA ENTROPÍA

Las restricciones fijan propiedades macro

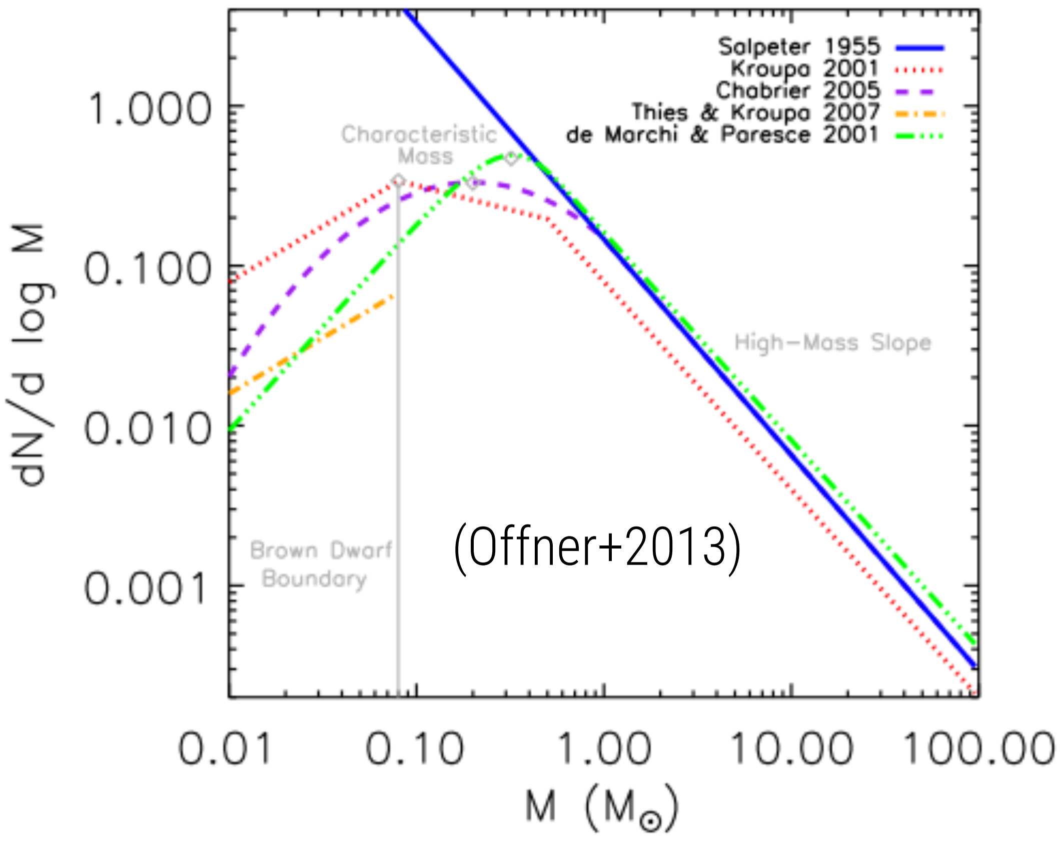
La información verificable define el conjunto de restricciones

$$\underbrace{L(P, P'; \mathcal{H})}_{\text{Información esperada ó Entropía, H(P)}} \equiv \int_{\mathcal{H}} \underbrace{P(\mathcal{H} \,|\, \mathcal{C})}_{\text{Distribución de Probabilidad a priori}} \times \underbrace{\log 1/P(\mathcal{H} \,|\, \mathcal{C})}_{\text{Información adquirida tras la observación}} \; ; \; \underbrace{\sum_{i=0} \lambda_i \, g_i(P, P'; \mathcal{H})}_{\text{Restricciones impuestas por la información contrastable}}$$

$$\max H : \underbrace{\frac{\partial L}{\partial P} - \frac{d}{d\mathcal{H}} \frac{\partial L}{\partial P'} + \sum_i \lambda_i \frac{\partial g_i}{\partial P} = 0}_{\text{Ecuación de Euler-Lagrange}}$$



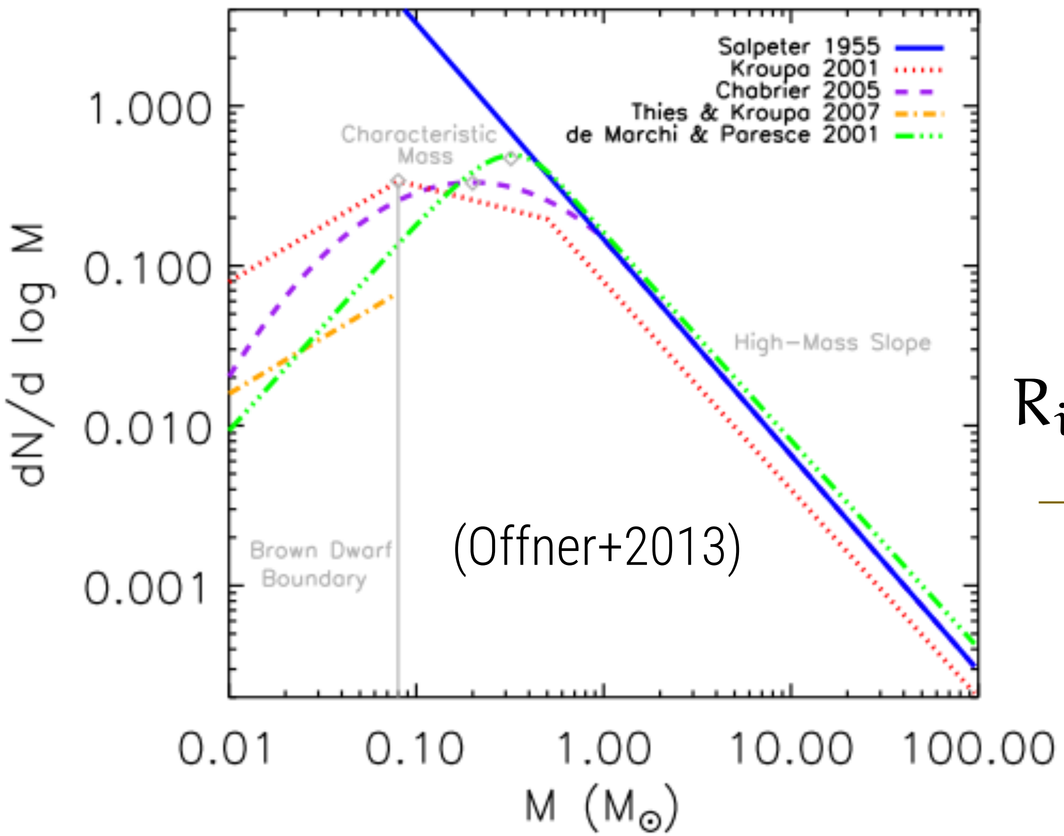
MÉTODO DE MÁXIMA ENTROPÍA – FUNCIÓN INICIAL DE MASA



$$P(m\,|\,\mathcal{C}) = \gamma\, m^{-\alpha}$$

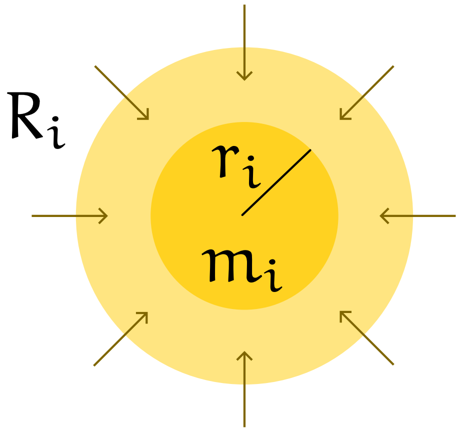
Vamos a recuperar la FIM de Salpeter usando simples suposiciones físicas y el método de Máxima Entropía.

MÉTODO DE MÁXIMA ENTROPÍA — FUNCIÓN INICIAL DE MASA



$$P(m\,|\,\mathcal{C}) = \gamma\, m^{-\alpha}$$

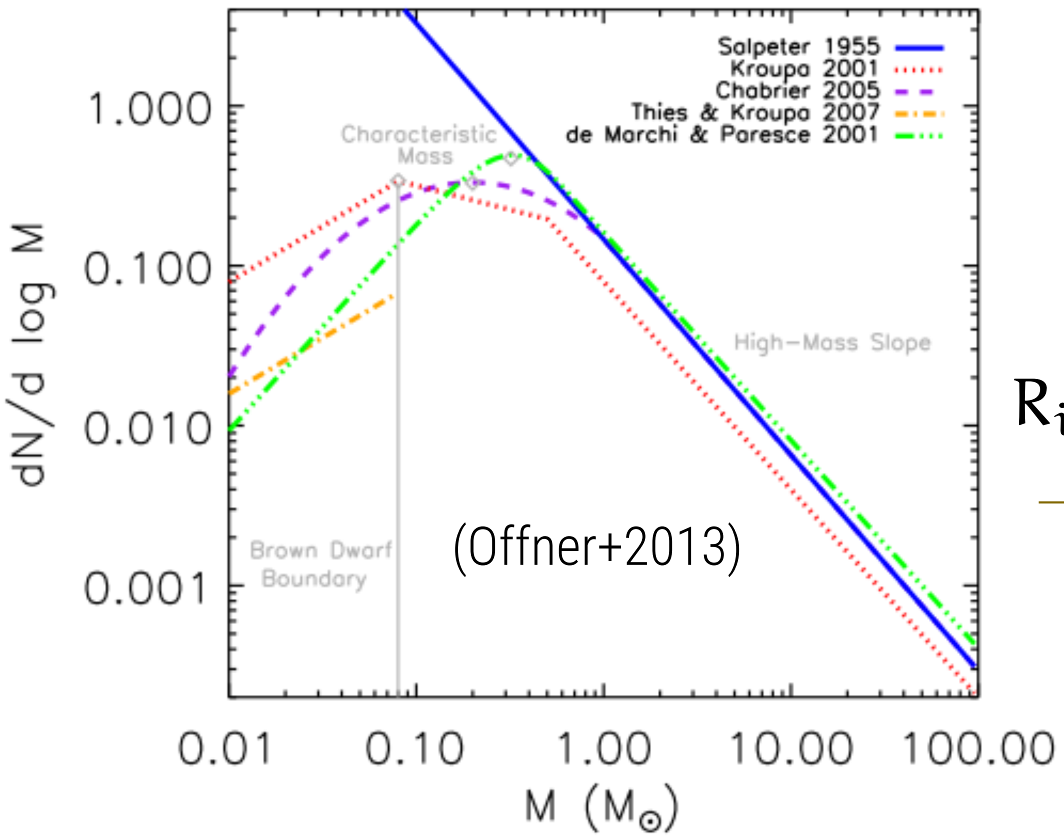
Vamos a recuperar la FIM de Salpeter usando simples suposiciones físicas y el método de Máxima Entropía.



$$R_i = \frac{m_i}{\tau_i^{\text{din}}}$$

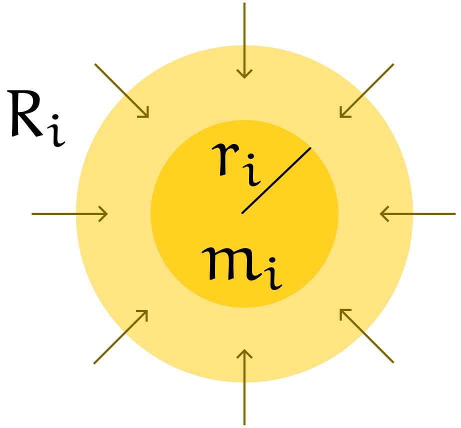
$$\tau^{\text{din}} \propto 1/\rho_i^{1/2}$$

MÉTODO DE MÁXIMA ENTROPÍA – FUNCIÓN INICIAL DE MASA



$$P(m\,|\,\mathcal{C}) = \gamma\, m^{-\alpha}$$

Vamos a recuperar la FIM de Salpeter usando simples suposiciones físicas y el método de Máxima Entropía.



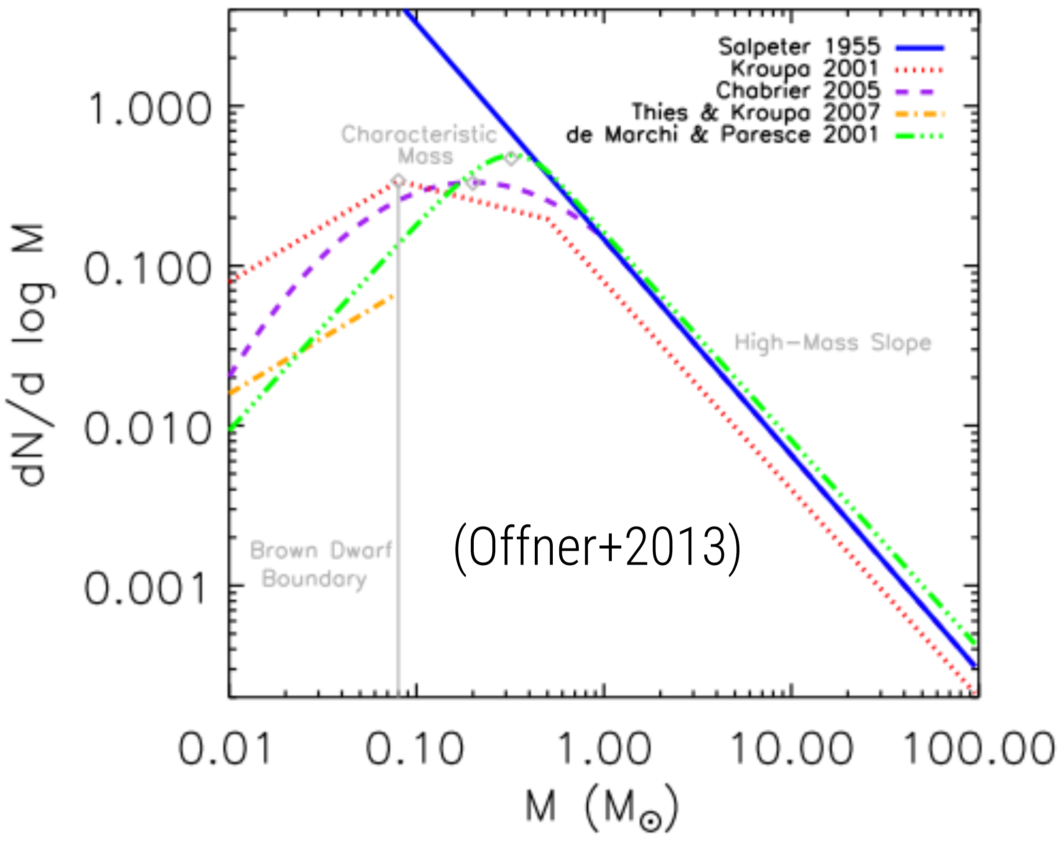
$$R_i = \frac{m_i}{\tau_i^{\mathrm{din}}}$$

$$\tau^{\mathrm{din}} \propto 1/\rho_i^{1/2}$$

$$R_i \propto (m_i\, r_i)^{3/2} \Longrightarrow \boxed{\rho_{R,i} \equiv R_i/V_i \propto (m_i/r_i)^{3/2}}$$

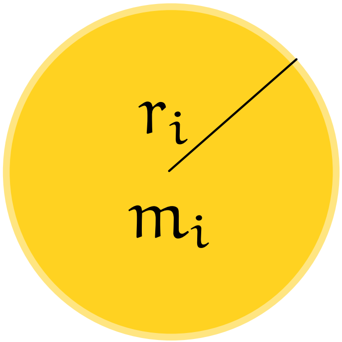
Queremos hallar el conjunto de $\{P_i\}$ tal que represente la DP de que una estrella haya alcanzado una masa m_i en la Secuencia Principal ($R_i = 0$).

MÉTODO DE MÁXIMA ENTROPÍA – FUNCIÓN INICIAL DE MASA



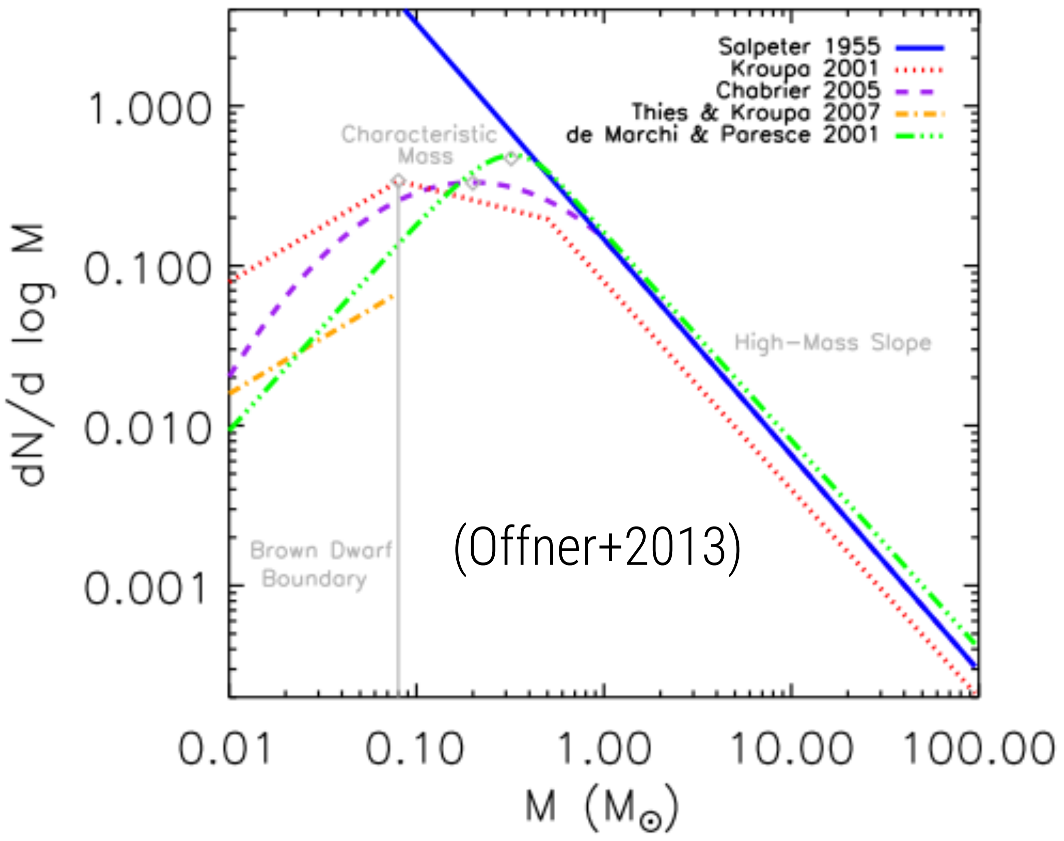
$$P(m|\mathcal{C}) = \gamma\, m^{-\alpha}$$

Vamos a recuperar la FIM de Salpeter usando simples suposiciones físicas y el método de Máxima Entropía.



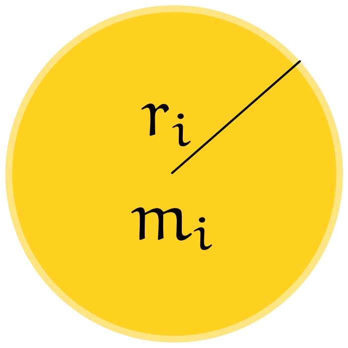
$$R_i = 0 \implies r_i = m_i^{4/5}$$

MÉTODO DE MÁXIMA ENTROPÍA – FUNCIÓN INICIAL DE MASA



$$P(m|\mathcal{C}) = \gamma\, m^{-\alpha}$$

Vamos a recuperar la FIM de Salpeter usando simples suposiciones físicas y el método de Máxima Entropía.



$$R_i = 0 \Longrightarrow r_i = m_i^{4/5}$$

$$\boxed{\rho_{R,i} = m_i^{1/5}}$$

Vamos a suponer que la densidad de Tasa de Acreción es una versión local de la Ley de Kennicutt-Schmidt:

$$\rho_{\mathrm{TFE}} \propto \rho_{\mathrm{gas}}^{3/2}$$

$$\Longrightarrow \boxed{\sum_i P_i \log m_i \propto \log \rho_{\mathrm{gas}} + \mathrm{const}}$$

MÉTODO DE MÁXIMA ENTROPÍA — FUNCIÓN INICIAL DE MASA

$$\underbrace{L(P_i, P_i', m_i) = \sum_i P_i \log 1/P_i}_{\text{Lagrangiano: Entropía}}$$

MÉTODO DE MÁXIMA ENTROPÍA – FUNCIÓN INICIAL DE MASA

$$\underbrace{L(P_i, P_i', m_i) = \sum_i P_i \log 1/P_i}_{\text{Lagrangiano: Entropía}}$$

$$\left.\begin{aligned} g_0 &= \sum_i P_i - 1 \\ g_1 &= \sum_i P_i \log m_i - \log \rho_{gas} \end{aligned}\right]$$

Restricciones basadas en la información contrastable

MÉTODO DE MÁXIMA ENTROPÍA – FUNCIÓN INICIAL DE MASA

$$\underbrace{L(P_i, P_i', m_i) = \sum_i P_i \log 1/P_i}_{\text{Lagrangiano: Entropía}}$$

$$\left.\begin{aligned} g_0 &= \sum_i P_i - 1 \\ g_1 &= \sum_i P_i \log m_i - \log \rho_{gas} \end{aligned}\right]$$

Restricciones basadas en la información contrastable

$$\underbrace{\frac{\partial L}{\partial P_i} - \frac{d}{dm_i}\frac{\partial L}{\partial P_i} + \sum_j \lambda_j \frac{\partial g_j}{\partial P_i}}_{\text{Ecuación de Euler-Lagrange}} = -\log P_i \underbrace{-1 + \lambda_0}_{\log \gamma} + \underbrace{\lambda_1}_{-\alpha} \log m_i = 0$$

MÉTODO DE MÁXIMA ENTROPÍA – FUNCIÓN INICIAL DE MASA

$$\underbrace{L(P_i, P_i', m_i) = \sum_i P_i \log 1/P_i}_{\text{Lagrangiano: Entropía}}$$

$$\left.\begin{aligned} g_0 &= \sum_i P_i - 1 \\ g_1 &= \sum_i P_i \log m_i - \log \rho_{gas} \end{aligned}\right]$$

Restricciones basadas en la información contrastable

$$\underbrace{\frac{\partial L}{\partial P_i} - \frac{d}{dm_i}\frac{\partial L}{\partial P_i} + \sum_j \lambda_j \frac{\partial g_j}{\partial P_i}}_{\text{Ecuación de Euler-Lagrange}} = -\log P_i \underbrace{- 1 + \lambda_0}_{\log \gamma} + \underbrace{\lambda_1}_{-\alpha} \log m_i = 0$$

$$\boxed{P_i = \gamma\, m_i^{-\alpha} \implies P(m \,|\, \mathcal{C}) = \gamma\, m^{-\alpha}}$$

MÉTODO DE MÁXIMA ENTROPÍA – FUNCIÓN INICIAL DE MASA

$$\underbrace{L(P_i, P_i', m_i) = \sum_i P_i \log 1/P_i}_{\text{Lagrangiano: Entropía}}$$

$$g_0 = \sum_i P_i - 1$$

$$g_1 = \sum_i P_i \log m_i - \log \rho_{gas}$$

Restricciones basadas en la información contrastable

$$\underbrace{\frac{\partial L}{\partial P_i} - \frac{d}{dm_i}\frac{\partial L}{\partial P_i} + \sum_j \lambda_j \frac{\partial g_j}{\partial P_i}}_{\text{Ecuación de Euler-Lagrange}} = -\log P_i \underbrace{-1 + \lambda_0}_{\log\gamma} + \underbrace{\lambda_1}_{-\alpha} \log m_i = 0$$

$$\boxed{P_i = \gamma\, m_i^{-\alpha} \implies P(m \mid \mathcal{C}) = \gamma\, m^{-\alpha}}$$

¡Hemos encontrado una explicación estocástica para la FIM, usando relaciones de escala de la Astrofísica y el método de Máxima Entropía!

RESUMEN

- La Distribución Previa debe asignarse de manera equilibrada.
- El método de Máxima Entropía proporciona la forma ideal de hacerlo.
- La Distribución Previa representa la mejor herramienta para muestrear la Verosimilitud.
- Sin embargo, la Inferencia Bayesiana presenta varias desventajas:

  Es computacionalmente costosa.

  No informa sobre qué tan bueno es el modelo propuesto.