MLB

$\beta_i | y_i, \underset{\sim}{x}_i \sim$ Modelo(η) = ¿?

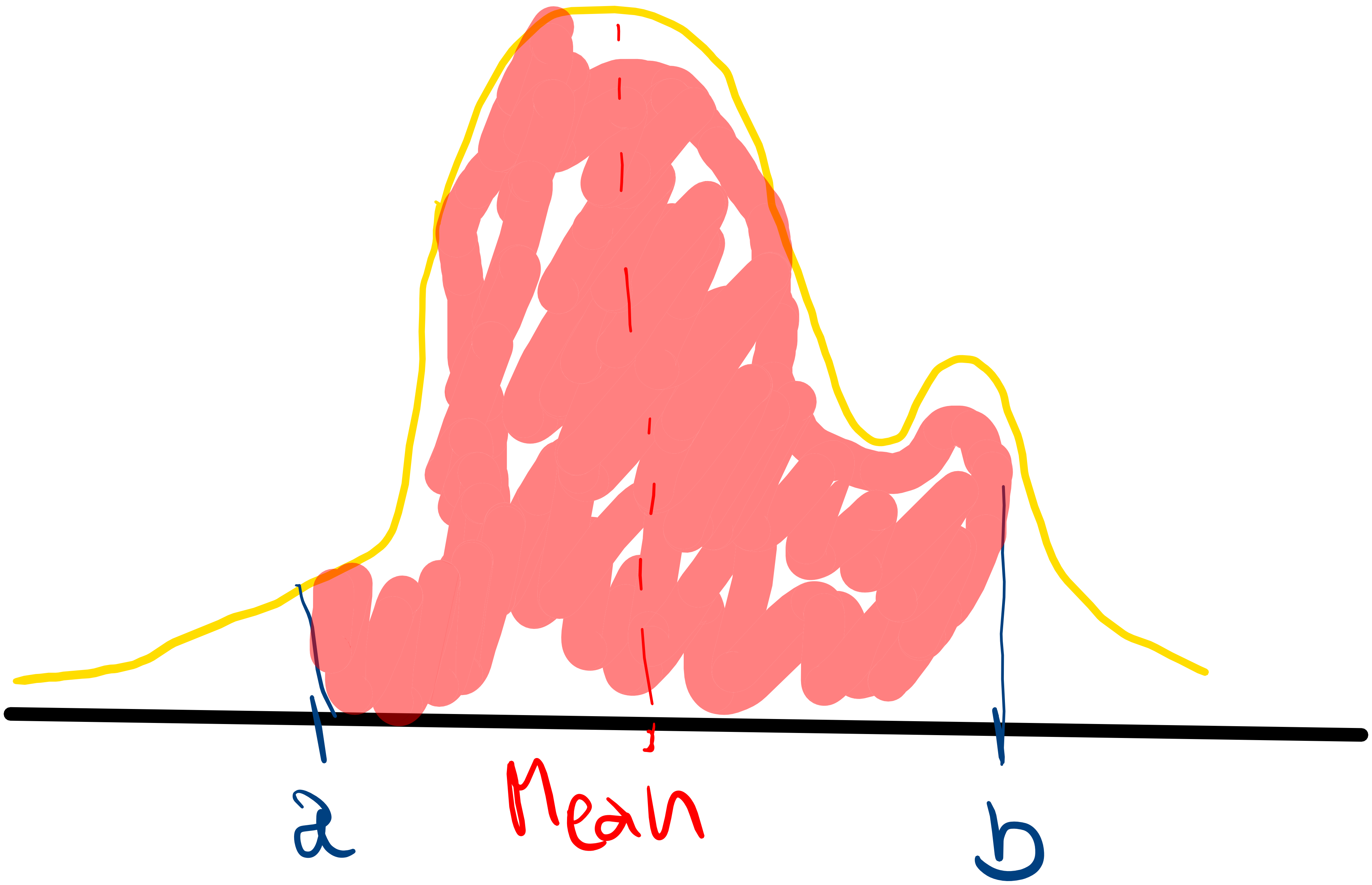
$\beta_i \sim$ Normal$(0, \sigma_i)$

$\mathbb{E}(\beta_i | y_i, \underset{-}{x}_i^T)$ = "mean"

HDI: high density interval
(a, b)

$P(a < \beta_i | y_i, x_i^T < b) = 0{,}97$ (97%)

- $\beta_i \mid y_i, x_i^T$



$$\beta_i \sim N(0,2)$$

$$\beta_i \mid y_i, x_i \sim \text{Modelo}(?)$$

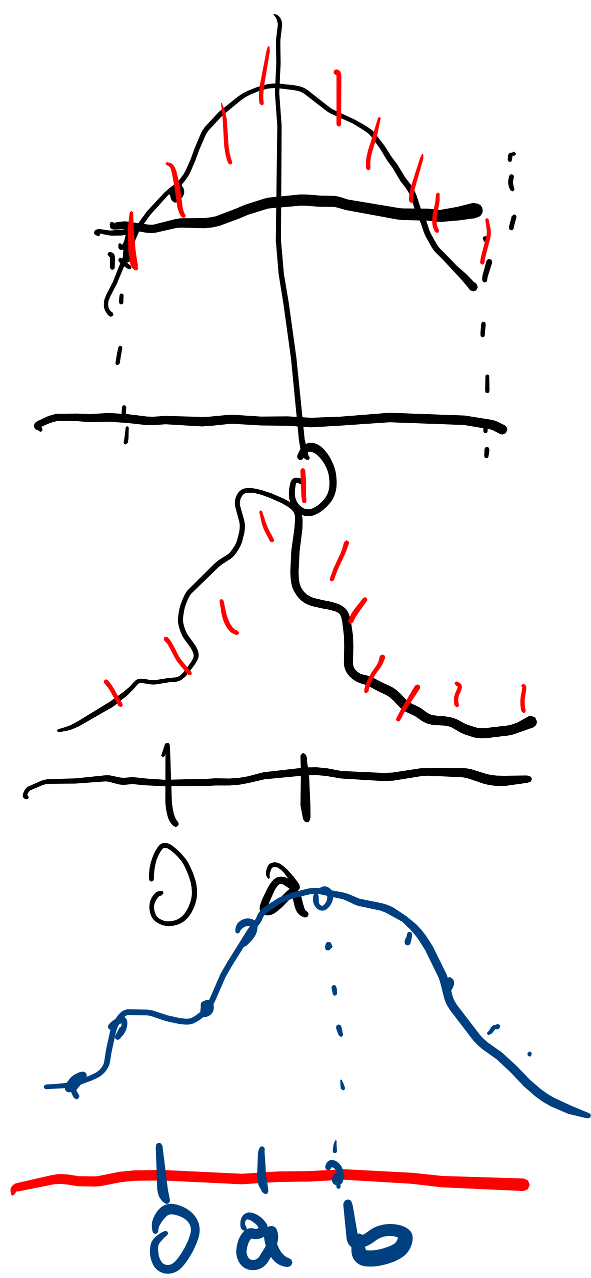
It.1: → Modelo 1: $\beta^1_i \mid y_i, x_i$

It.2: Modelo 2: $\beta^2_i \mid y_i, x_i$

~~converger~~

5000 ...

It.5000 Modelo 5000 $\beta_i \mid y_i x_i$ 5000



→ $t \geq 4000$, Existe convergencia

→ $\left\{\bar{\beta}^t\right\}_{t=4000}^{5000}$, $\bar{\bar{\beta}} = \frac{1}{1000} \sum_{i=1}^{1000} \bar{\beta}^t$

$P(\beta^t \mid \beta^{t-1})$

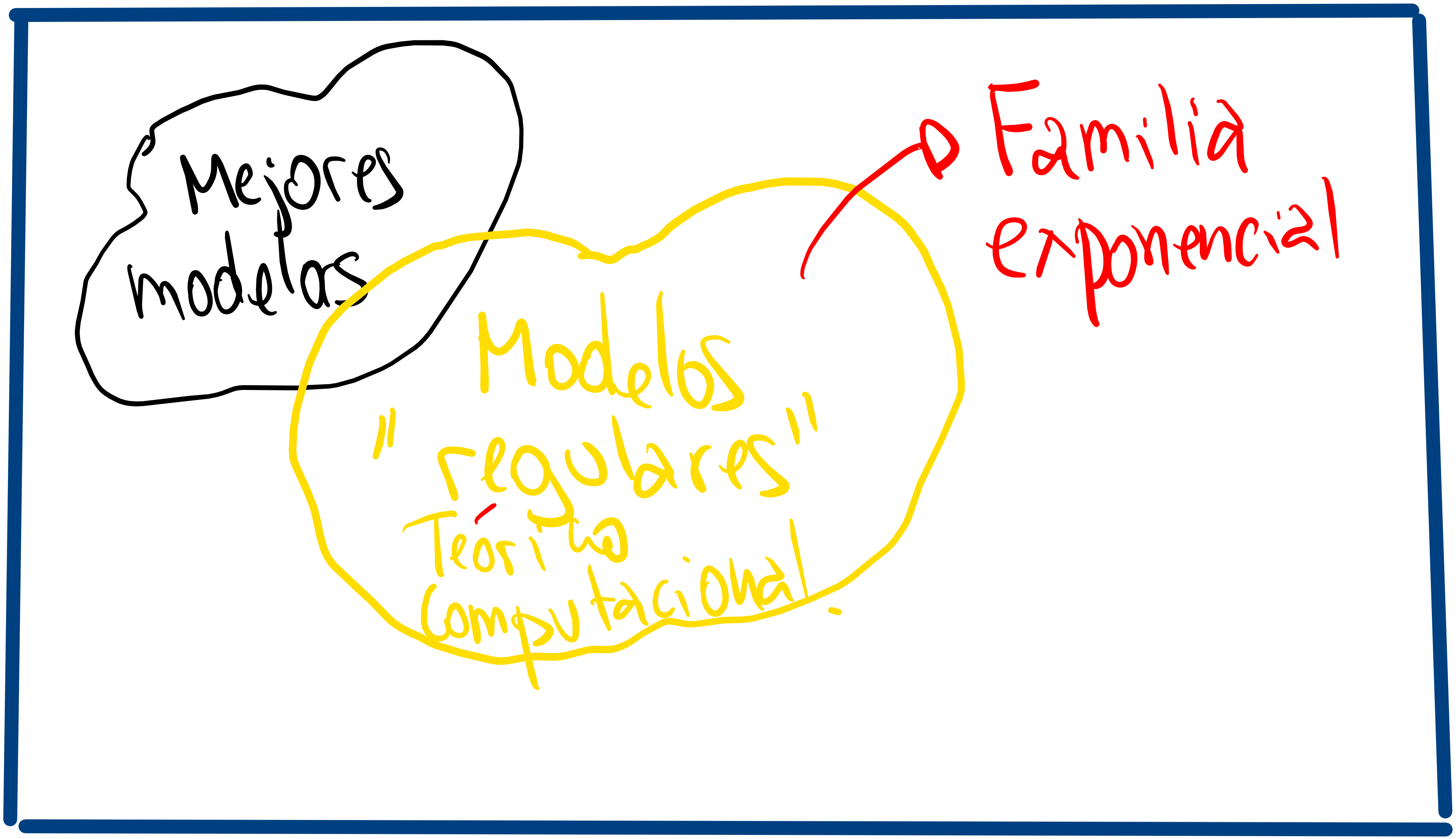
> Gibbs sampling

> Variational sampling. ✓

Clase 3

- Familia exponencial
- Reducción de X por estadísticas.
- Información de Fisher.

$\mathcal{P} = \{ P : P$ cumple alguna prop $\}$. Modelos



Parte I

Familia exponencial

Familia exponencial

$$p_\theta(x) = h(x)\, e^{\langle \eta(\theta), T(x)\rangle - A(\theta)}$$

$$\langle \eta(\theta), T(x)\rangle = \sum_{i=1}^{d} \eta_i(\theta)\, T_i(x) \to \eta(\theta)\, T(x).$$

Forma especial

$$p_\theta(x) = a(\theta)\, h(x)\, e^{\langle \eta(\theta), T(x)\rangle} ; \quad a(\theta) = e^{-A(\theta)}$$

Ejemplo: modelo logístico

<u>Continuo</u>: $P_\theta(X \in A)$; μ : Lebesgue

$$\mu(A) = 0 \Rightarrow P_\theta(X \in A) = 0$$

<u>Discreto:</u> $P'_\theta(X \in B)$; μ : conteo

%,

$$P_\theta(X \in A) = \int_A a(\theta)\, h(x)\, e^{\langle \theta, T(x) \rangle}\, d\mu(x) < \infty$$

Propiedades del espacio paramétrico. 2.1

Si $\eta(\theta)=\theta$, entonces $p_\theta(x)=a(\theta)h(x)e^{\langle\theta, T(x)\rangle}$.

Θ es el espacio paramétrico

$$\Theta=\left\{\theta : \int h(x)e^{\langle\theta, T(x)\rangle}\,d\mu(x)<\infty\right\}$$

$\eta(\theta)=\theta.$

Resultado: Θ es convexo.

Implicancias

> El EMV es único

> Los algoritmos serán eficientes.

> Estabilidad en regularización y penalización

> Asegura que siempre habrá convergencia.

Ejemplo: modelo logístico

$\theta \to \beta$

$$P_\beta(y) = \exp\left\{ y \cdot x^T\beta - \log(1 + e^{x^T\beta}) \right\}$$

$T(y) = y$; $\eta(\beta) = x^T\beta$; $A(\beta) = \log(1 + e^{x^T\beta})$

$h(y) = 1$.

→ En general, todo MLG.

- <u>Propiedad: cambio en el orden de integrada y derivada</u>

$$P_\theta(X \in A) = a(\theta) \int_{\underline{\ \ }} h(x)\, e^{\langle\theta, T(x)\rangle} d\mu(x) < \infty$$

$$\ell(x) = g(x)\, h(x)$$

X, x, Y, y

$$\mathbb{E}(|\ell(X)|) = a(\theta) \int |\ell(x)|\, e^{\langle\theta, T(x)\rangle} d\mu(x$$

Caso 1: $\ell(x) = 1$. o $\ell(x) = g(x)\, h(x)$

<u>caso 2:</u>

Aplicación al modelo Logístico

Para alguna $\ell(x)$ (1)

$$\Theta_\ell \subseteq \Theta$$

$$m(\theta) = \int \ell(x) e^{\langle \theta, T(x) \rangle} d\mu(x)$$

- Es continua $(\theta \in \Theta_\ell)$
- $m^{(r)}(\theta)$ existe $\forall r \in \mathbb{N}$.

$$\theta = (\theta_1, \dots, \theta_p)$$

$$\frac{\partial m(\theta)}{\partial \theta_l} = \int x_i \, f(x) \, e^{\langle \theta, x \rangle} \, d\mu(x) \quad //$$

Condiciones de regularidad Sigla

$\downarrow$ sobre

$$\mathcal{P}_{\Theta} = \{ P_\theta : \theta \in \Theta \},$$

Aplicación de la familia exponencial a Data Science

•

Ronald Fisher//

$(\mathcal{X}, \mathcal{G}, \mathcal{P})$

Parte 2

suficiencia
Ancilaridad
Completitud

$T: \mathcal{X} \to \mathcal{Y}$

$\dim(T(x)) \leq \dim(x)$//

- T que resuma sin perder "información"
- A que no tengo información sobre θ.

Propiedad de una estadística suficiente

- $P_\theta(X \in A \mid T(x) = t) = P'(X \in A, t)$
- $P_\theta(X = x \mid T(x) = t) = P_t(X = x)$
- $T(x)$ tiene toda info. de x sobre θ.
- $P_t(x)$ es invariante con respecto a θ

Inferencia

P_θ que $X = x_D$ desconocida

$$P_\theta(X \in A \mid T(x) = t) = P'_t(X \in A)$$

$T(x) = t$ conocida

Simular muestras de X $(x'_1, \dots, x'_K)$ desde $P'_t(X \in A)$ // $(p_1, p_2, \dots, p_K)$

Bernoulli

$x \in \{0,1\}^n$, $x_0 = (0,1,0,0,\dots,1)$

$$P(X = x \mid \underline{T(x) = t}) = \frac{1}{\binom{m}{t}} \quad (\not\exists x)$$

modelo uniforme

$\underline{x_1 \mid T=t, \dots, x_k \mid T=t},$

$T' = \frac{1}{m}\sum x_i$

$T'' = g(T)$

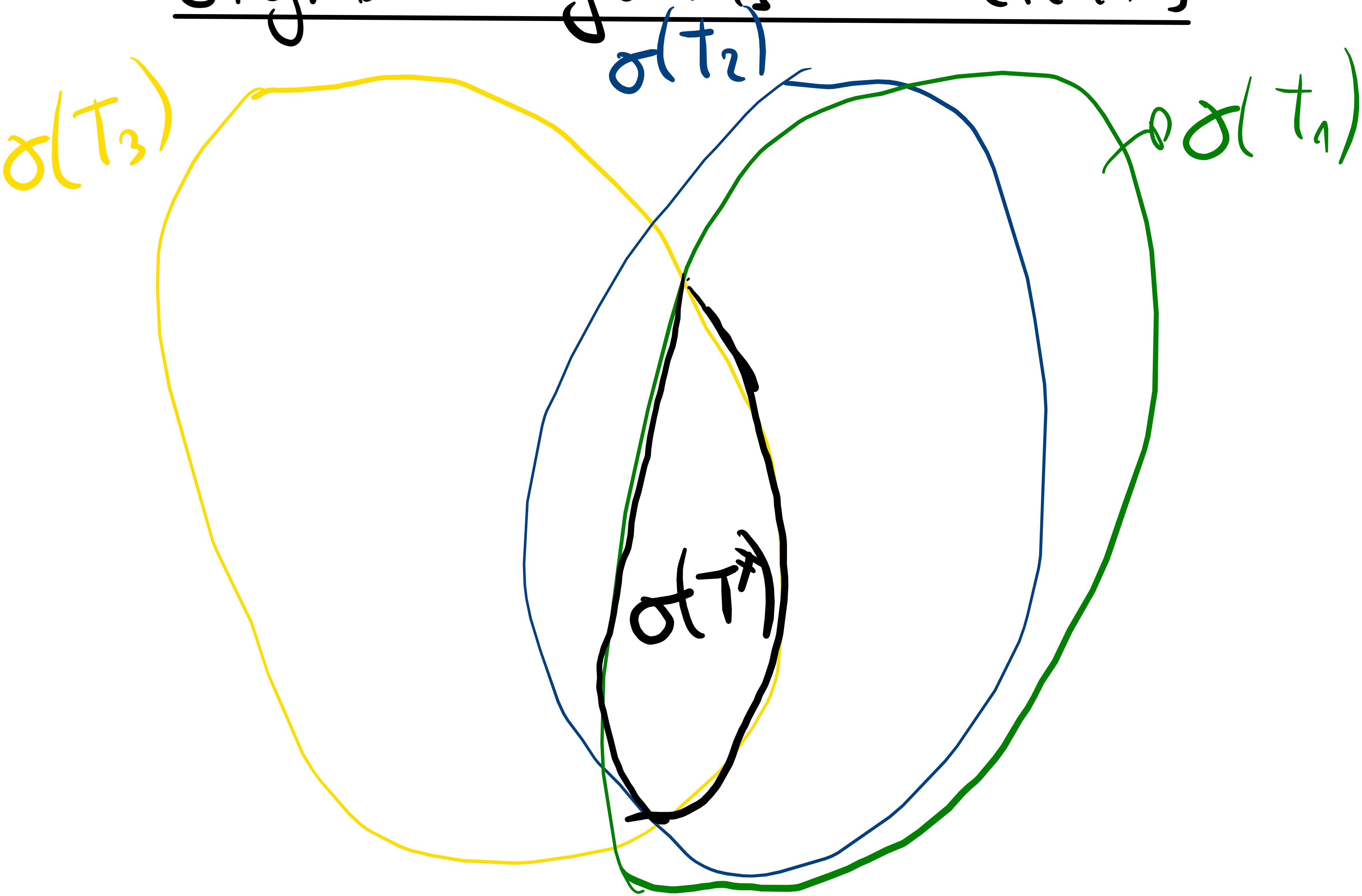
$\sigma(T) \neq \sigma(T') \neq \sigma(T'')$

Estadística suficiente minimal

> T^* es suf. minimal si es función de toda estadística suficiente T.
>
> $\forall T$ suficiente $\exists h: T^* = h(T)$

> $\sigma(T^*) \subseteq \sigma(T) \subseteq \sigma(X)$

Sigma algebras suficientes



Estadística completa e ancilar.

A: $\mathcal{X} \to \mathcal{A}$ es ancilar para θ, si no tiene información sobre θ.

$$P_\theta(A \in Z) = P''(A \in Z) \quad \forall \theta \in \Theta$$

C: $\mathcal{X} \to \mathcal{C}$ es llamado de completo si C es suficiente y no tiene información ancilar.

$\big|_{(\mathcal{A}, \mathcal{L}, P_A)}$ $Z \in \mathcal{L}$.

Teorema de Basu

T es completa y A es ancilar $\Rightarrow$

$$T \perp\!\!\!\perp A$$

$\rightarrow$ I.C.

Teorema de la factorización

Ejemplo de aplicación

La estadística

$$T(y) = \sum_{i=1}^{n} y_i x_i^T \quad \text{es}$$

suficiente para P_β //.

Parte 3

Información de Fisher

Curvatura $L_x(\theta)$ o $\log L_x(\theta)$,,

$$\mathcal{P}_{\Theta} = \{P_\theta : \theta \in \Theta\} \qquad p_\theta(x) = \frac{dP_\theta}{d\mu(x)}$$

$$\theta = (\theta_1, \dots, \theta_p)$$

Definición 2.4 Función Score e Información de Fisher

Suponga que se cumplen las condiciones de regularidad de la Información de Fisher. El vector score se define como $\frac{\partial \log p_\theta(X)}{\partial \theta_i}$ para $i = 1, \dots, d$. La Información de Fisher $I_X(\theta)$ es la matriz de covarianza del vector score; es decir,

$$I_X(\theta)_{ij} = C_\theta\left(\frac{\partial \log p_\theta(X)}{\partial \theta_i}, \frac{\partial \log p_\theta(X)}{\partial \theta_j}\right). \tag{6}$$

Se tiene que el valor esperado del score es cero. En este caso, si podemos diferenciar dos veces dentro del signo integral (como en familias exponenciales; cf. Teorema 1), entonces hay una fórmula alternativa para la Información de Fisher:

$$I_X(\theta)_{ij} = -\mathbb{E}_\theta\left\{\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p_\theta(X)\right\}.$$

$$i = j$$

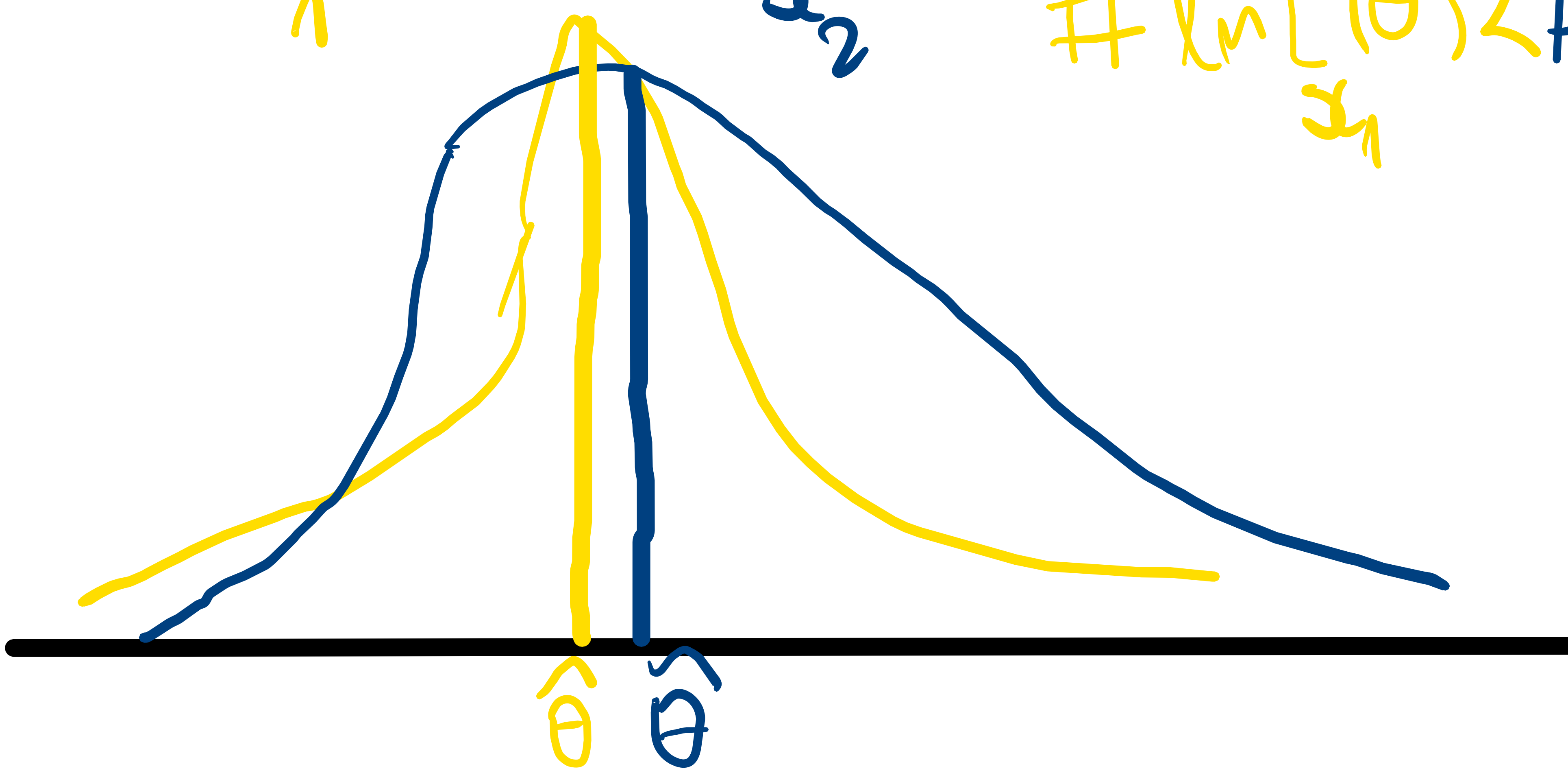
$$I_x(\theta_i) = I_x(\theta)_{ii}$$

$$I_x(\theta) = \left(I_x(\theta_1), \ldots, I_x(\theta_p) \right)$$

$I_X(\theta)$ es una medida la curvatura de $\ln L(\theta)$

$$\ln L_{x_1}(\theta) = \ln L_{x_2}(\theta)$$

$$\# \ln L_{x_1}(\theta) < \# \ln L_{x_2}(\theta)$$



Teorema

$$I_X(\theta) = I_T(\theta)$$

si T es suficiente.

Interpretación geométrica

Resultado

Aplicaciones a Data Science