Estimación de densidades mediante modelos de mezcla controlados por el proceso de Dirichlet

Estudiantes:

Manuel Hernández Banadik
Mario Sierra Golombievski

Tutor:

Ignacio Alvarez-Castro

2 de julio de 2019

Plan

Inferencia Bayesiana

Supongamos que se observan los datos $\mathbf{y} = (y_1, \ldots, y_n)$.

- $\mathbf{y} \sim f(\mathbf{y}|\theta)$ es lo que se denomina **modelo para los datos**

Inferencia Bayesiana

Supongamos que se observan los datos $\mathbf{y} = (y_1, \ldots, y_n)$.

- $\mathbf{y} \sim f(\mathbf{y}|\theta)$ es lo que se denomina **modelo para los datos**
- $f(\theta)$ es la **distribución previa** para el parámetro

Inferencia Bayesiana

Supongamos que se observan los datos $\mathbf{y} = (y_1, \ldots, y_n)$.

- $\mathbf{y} \sim f(\mathbf{y}|\theta)$ es lo que se denomina **modelo para los datos**
- $f(\theta)$ es la **distribución previa** para el parámetro
- Por el Teorema de Bayes se obtiene la **distribución posterior**:

$$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)f(\theta)}{f(\mathbf{y})} = \frac{f(\mathbf{y}|\theta)f(\theta)}{\int f(\mathbf{y}|\theta)f(\theta)d\theta} \propto f(\mathbf{y}|\theta)f(\theta)$$

Inferencia Bayesiana

Supongamos que se observan los datos $\mathbf{y} = (y_1, \ldots, y_n)$.

- $\mathbf{y} \sim f(\mathbf{y}|\theta)$ es lo que se denomina **modelo para los datos**
- $f(\theta)$ es la **distribución previa** para el parámetro
- Por el Teorema de Bayes se obtiene la **distribución posterior**:

$$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)f(\theta)}{f(\mathbf{y})} = \frac{f(\mathbf{y}|\theta)f(\theta)}{\int f(\mathbf{y}|\theta)f(\theta)d\theta} \propto f(\mathbf{y}|\theta)f(\theta)$$

En general, esta no se puede calcular analíticamente, siendo necesario recurrir a métodos computacionales para estimarla.

Inferencia Bayesiana

Supongamos que se observan los datos $\mathbf{y} = (y_1, \ldots, y_n)$.

- $\mathbf{y} \sim f(\mathbf{y}|\theta)$ es lo que se denomina **modelo para los datos**
- $f(\theta)$ es la **distribución previa** para el parámetro
- Por el Teorema de Bayes se obtiene la **distribución posterior**:

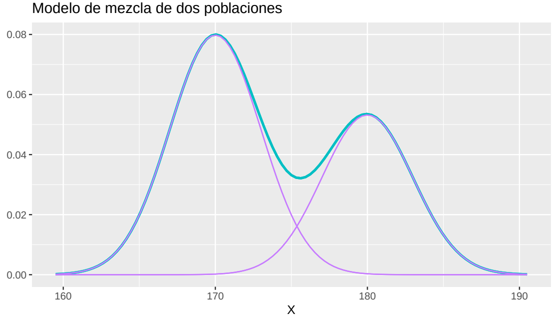
$$f(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)f(\theta)}{f(\mathbf{y})} = \frac{f(\mathbf{y}|\theta)f(\theta)}{\int f(\mathbf{y}|\theta)f(\theta)d\theta} \propto f(\mathbf{y}|\theta)f(\theta)$$

En general, esta no se puede calcular analíticamente, siendo necesario recurrir a métodos computacionales para estimarla.

Cuando $f(\theta)$ y $f(\theta|\mathbf{y})$ pertenecen a la misma familia, se dice que la distribución previa es conjugada natural.

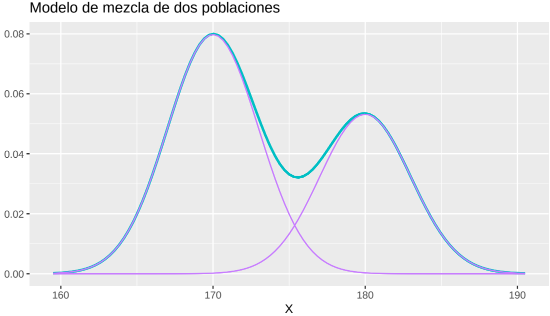
Modelos de mezcla

$$f(x) = 0.6 \times \phi(x; 170, 1) + 0.4 \times \phi(x; 180, 1)$$



Modelos de mezcla

$$f(x) = 0.6 \times \phi(x; 170, 1) + 0.4 \times \phi(x; 180, 1)$$

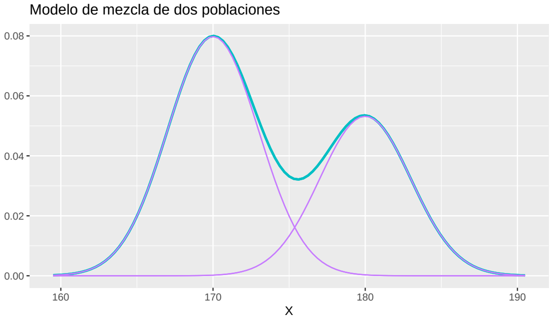


Un modelo de mezcla finito tiene la forma:

$$f(x) = \pi_1 f_1(x|\theta_1) + \cdots + \pi_H f_H(x|\theta_H)$$

Modelos de mezcla

$$f(x) = 0.6 \times \phi(x; 170, 1) + 0.4 \times \phi(x; 180, 1)$$



Un modelo de mezcla finito tiene la forma:

$$f(x) = \pi_1 f_1(x|\theta_1) + \cdots + \pi_H f_H(x|\theta_H)$$

En general

$$f(x) = \int_\Theta f(x|\theta) d\mathrm{P}(\theta)$$

Modelo Dirichlet-Multinomial

Sean $X_1, \ldots, X_n$ i.i.d. con $Rec(X_i) = \{1, 2, \ldots, k\}$ tales que $\mathbb{P}(X_i = j) = \pi_j$ para $j = 1, 2, \ldots, k$.

Si $Y_j := \sum_{i=1}^{n} \mathbb{1}_{\{X_i = j\}}$ para $j = 1, \ldots, k \Rightarrow (Y_1, \ldots, Y_k) \sim Multinomial(\pi_1, \ldots, \pi_k)$.

Modelo Dirichlet-Multinomial

Sean $X_1, \ldots, X_n$ i.i.d. con $Rec(X_i) = \{1, 2, \ldots, k\}$ tales que $\mathbb{P}(X_i = j) = \pi_j$ para $j = 1, 2, \ldots, k$.

Si $Y_j := \sum_{i=1}^{n} \mathbb{1}_{\{X_i = j\}}$ para $j = 1, \ldots, k \Rightarrow (Y_1, \ldots, Y_k) \sim Multinomial(\pi_1, \ldots, \pi_k)$.

Distribución de Dirichlet

Su soporte es el conjunto de los vectores de probabilidad de dimensión k, $(\mathscr{P}_k)$.

$(\pi_1, \ldots, \pi_k) \sim Dirichlet(a_1, \ldots, a_k)$, $(a_i > 0 \forall i)$ si:

$$f(\pi_1, \ldots, \pi_k | a_1, \ldots, a_k) = \frac{\Gamma\left(\sum_{j=1}^{k} a_j\right)}{\prod_{j=1}^{k} \Gamma(a_j)} \prod_{j=1}^{k} \pi_j^{a_j - 1} \mathbb{1}_{\{\mathscr{P}_k\}}(\pi_1, \ldots, \pi_k)$$

Modelo Dirichlet-Multinomial

Sean $X_1, \ldots, X_n$ i.i.d. con $Rec(X_i) = \{1, 2, \ldots, k\}$ tales que $\mathbb{P}(X_i = j) = \pi_j$ para $j = 1, 2, \ldots, k$.

Si $Y_j := \sum_{i=1}^{n} \mathbb{1}_{\{X_i = j\}}$ para $j = 1, \ldots, k \Rightarrow (Y_1, \ldots, Y_k) \sim \mathit{Multinomial}(\pi_1, \ldots, \pi_k)$.

Distribución de Dirichlet

Su soporte es el conjunto de los vectores de probabilidad de dimensión k, $(\mathscr{P}_k)$.

$(\pi_1, \ldots, \pi_k) \sim \mathit{Dirichlet}(a_1, \ldots, a_k)$, $(a_i > 0 \forall i)$ si:

$$f(\pi_1, \ldots, \pi_k | a_1, \ldots, a_k) = \frac{\Gamma\left(\sum_{j=1}^{k} a_j\right)}{\prod_{j=1}^{k} \Gamma(a_j)} \prod_{j=1}^{k} \pi_j^{a_j - 1} \mathbb{1}_{\{\mathscr{P}_k\}}(\pi_1, \ldots, \pi_k)$$

- Todas sus marginales π_j se distribuyen Beta

Modelo Dirichlet-Multinomial

Sean $X_1, \ldots, X_n$ i.i.d. con $Rec(X_i) = \{1, 2, \ldots, k\}$ tales que $\mathbb{P}(X_i = j) = \pi_j$ para $j = 1, 2, \ldots, k$.

Si $Y_j := \sum_{i=1}^{n} \mathbb{1}_{\{X_i = j\}}$ para $j = 1, \ldots, k \Rightarrow (Y_1, \ldots, Y_k) \sim Multinomial(\pi_1, \ldots, \pi_k)$.

Distribución de Dirichlet

Su soporte es el conjunto de los vectores de probabilidad de dimensión k, $(\mathscr{P}_k)$.

$(\pi_1, \ldots, \pi_k) \sim Dirichlet(a_1, \ldots, a_k)$, $(a_i > 0 \forall i)$ si:

$$f(\pi_1, \ldots, \pi_k | a_1, \ldots, a_k) = \frac{\Gamma\left(\sum_{j=1}^{k} a_j\right)}{\prod_{j=1}^{k} \Gamma(a_j)} \prod_{j=1}^{k} \pi_j^{a_j - 1} \mathbb{1}_{\{\mathscr{P}_k\}}(\pi_1, \ldots, \pi_k)$$

- Todas sus marginales π_j se distribuyen Beta
- Es un modelo conjugado. Si se observa $\mathbf{y}_1, \ldots, \mathbf{y}_N$ donde $\mathbf{y}_i = (y_{1,i}, \ldots, y_{k,i})$

Modelo Dirichlet-Multinomial

Sean $X_1, \ldots, X_n$ i.i.d. con $Rec(X_i) = \{1, 2, \ldots, k\}$ tales que $\mathbb{P}(X_i = j) = \pi_j$ para $j = 1, 2, \ldots, k$.

Si $Y_j := \sum_{i=1}^{n} \mathbb{1}_{\{X_i = j\}}$ para $j = 1, \ldots, k \Rightarrow (Y_1, \ldots, Y_k) \sim \textit{Multinomial}(\pi_1, \ldots, \pi_k)$.

Distribución de Dirichlet

Su soporte es el conjunto de los vectores de probabilidad de dimensión k, $(\mathscr{P}_k)$.

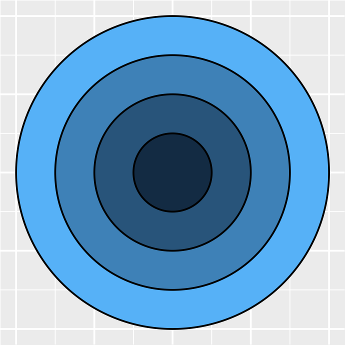
$(\pi_1, \ldots, \pi_k) \sim \textit{Dirichlet}(a_1, \ldots, a_k)$, $(a_i > 0 \forall i)$ si:

$$f(\pi_1, \ldots, \pi_k | a_1, \ldots, a_k) = \frac{\Gamma\left(\sum_{j=1}^{k} a_j\right)}{\prod_{j=1}^{k} \Gamma(a_j)} \prod_{j=1}^{k} \pi_j^{a_j - 1} \mathbb{1}_{\{\mathscr{P}_k\}}(\pi_1, \ldots, \pi_k)$$

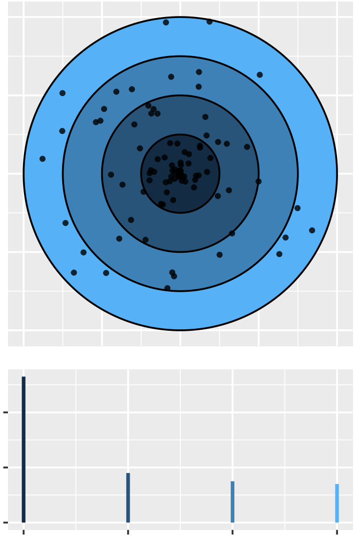
- Todas sus marginales π_j se distribuyen Beta
- Es un modelo conjugado. Si se observa $\mathbf{y}_1, \ldots, \mathbf{y}_N$ donde $\mathbf{y}_i = (y_{1,i}, \ldots, y_{k,i})$ entonces

$$(\pi_1, \ldots, \pi_k) | \mathbf{y}_1, \ldots, \mathbf{y}_N \sim \textit{Dirichlet}(a_1 + \sum_{i=1}^{N} y_{1,i}, \ldots, a_k + \sum_{i=1}^{N} y_{k,i})$$

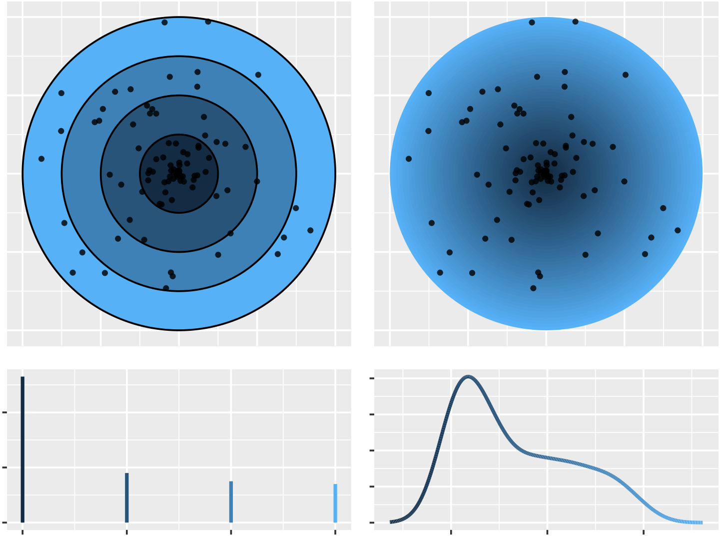
Proceso de Dirichlet



Proceso de Dirichlet



Proceso de Dirichlet



Medida de Probabilidad Aleatoria

Dado un espacio probabilizable $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ y un espacio de probabilidad $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$, definimos medida de probabilidad aleatoria a la función:

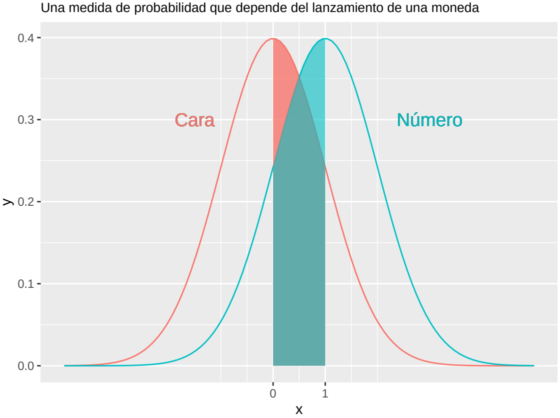
$$\mathcal{P} : \Omega \times \mathcal{B}(\mathcal{X}) \to [0, 1]$$

tal que:

- Fijado $\omega \in \Omega$, para todo $B \in \mathcal{B}(\mathcal{X})$, el mapeo $B \mapsto \mathcal{P}(\omega, B)$ es una medida de probabilidad sobre el espacio probabilizable $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$
- Fijado $B \in \mathcal{B}(\mathcal{X})$, para todo $\omega \in \Omega$, el mapeo $\omega \mapsto \mathcal{P}(\omega, B)$ es una variable aleatoria que toma valores en $[0, 1]$.

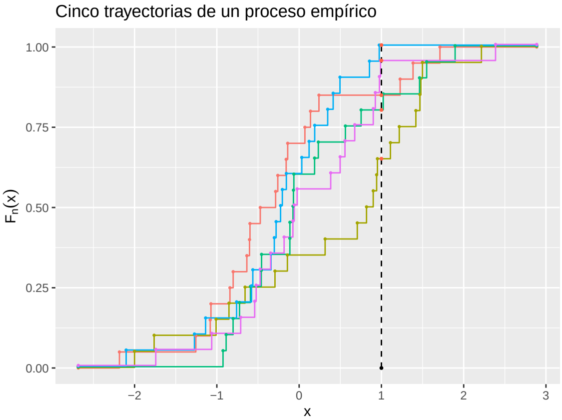
Medida de Probabilidad Aleatoria

- Fijado $\omega \in \Omega$, para todo $B \in \mathcal{B}(\mathcal{X})$, el mapeo $B \mapsto \mathcal{P}(\omega, B)$ es una medida de probabilidad sobre el espacio probabilizable $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$
- Fijado $B \in \mathcal{B}(\mathcal{X})$, para todo $\omega \in \Omega$, el mapeo $\omega \mapsto \mathcal{P}(\omega, B)$ es una variable aleatoria que toma valores en $[0, 1]$.



Medida de Probabilidad Aleatoria

- Fijado $\omega \in \Omega$, para todo $B \in \mathcal{B}(\mathcal{X})$, el mapeo $B \mapsto \mathcal{P}(\omega, B)$ es una medida de probabilidad sobre el espacio probabilizable $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$
- Fijado $B \in \mathcal{B}(\mathcal{X})$, para todo $\omega \in \Omega$, el mapeo $\omega \mapsto \mathcal{P}(\omega, B)$ es una variable aleatoria que toma valores en $[0, 1]$.



Proceso de Dirichlet

Sea $(\Omega, \mathscr{B}(\mathbb{R}), \mathbb{P})$ un espacio de probabilidad, $\mathcal{P}$ una medida de probabilidad aleatoria sobre $(\Omega, \mathscr{B}(\mathbb{R}), \mathbb{P})$.

Proceso de Dirichlet

Sea $(\Omega, \mathscr{B}(\mathbb{R}), \mathbb{P})$ un espacio de probabilidad, $\mathcal{P}$ una medida de probabilidad aleatoria sobre $(\Omega, \mathscr{B}(\mathbb{R}), \mathbb{P})$.

Diremos que $\mathcal{P}$ está definida por un proceso de Dirichlet de parámetros α y G_0 si para cualquier partición medible de $\mathbb{R}$ de la forma $\mathcal{A} = (A_1, \ldots, A_k)$, se cumple que:

$$\mathcal{P}(\mathcal{A}) = \Big(\mathcal{P}(A_1), \mathcal{P}(A_2), \ldots, \mathcal{P}(A_k)\Big) \sim \mathit{Dirichlet}(\alpha G_0(A_1), \ldots, \alpha G_0(A_k))$$

Proceso de Dirichlet

Sea $(\Omega, \mathscr{B}(\mathbb{R}), \mathbb{P})$ un espacio de probabilidad, $\mathcal{P}$ una medida de probabilidad aleatoria sobre $(\Omega, \mathscr{B}(\mathbb{R}), \mathbb{P})$.

Diremos que $\mathcal{P}$ está definida por un proceso de Dirichlet de parámetros α y G_0 si para cualquier partición medible de $\mathbb{R}$ de la forma $\mathcal{A} = (A_1, \ldots, A_k)$, se cumple que:

$$\mathcal{P}(\mathcal{A}) = \Big(\mathcal{P}(A_1), \mathcal{P}(A_2), \ldots, \mathcal{P}(A_k)\Big) \sim Dirichlet(\alpha G_0(A_1), \ldots, \alpha G_0(A_k))$$

y lo notaremos: $\mathcal{P} \sim DP(\alpha, G_0)$, donde $\alpha > 0$ es un parámetro de precisión y G_0 es una medida de probabilidad base.

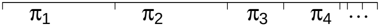
Stick-Breaking

En [Sethuraman, 1994] se prueba que si $\mathcal{P}$ es un proceso de Dirichlet con parámetros (α, G_0), entonces, se puede escribir de la forma:

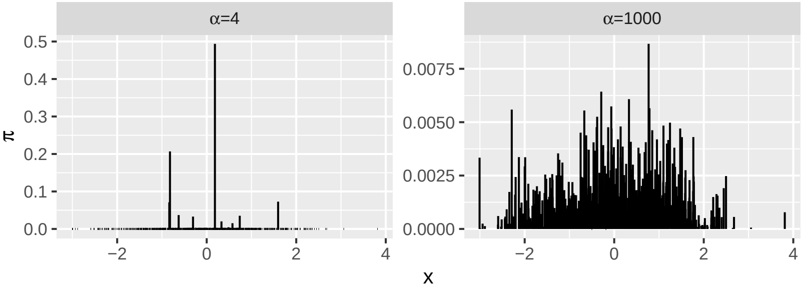
$$\mathcal{P} = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k} \tag{1}$$

donde δ_x es la *Delta de Dirac*

- $\theta_1, \theta_2, \cdots \overset{ind}{\sim} G_0$
- $\pi_1, \pi_2, \cdots \sim Stick(\alpha)$



Dos realizaciones de stick breaking



Estimación de una distribución

Modelo para los datos:

$$X_1, \ldots, X_n \overset{ind}{\sim} \mathcal{P}$$

(Una realización de una M.P.A.)

Estimación de una distribución

Modelo para los datos:

$$X_1, \ldots, X_n \overset{ind}{\sim} \mathcal{P}$$

(Una realización de una M.P.A.)

Distribución previa:

$$\mathcal{P} \sim DP(\alpha, G_0)$$

Estimación de una distribución

Modelo para los datos:

$$X_1, \ldots, X_n \overset{ind}{\sim} \mathcal{P}$$

(Una realización de una M.P.A.)

Distribución previa:

$$\mathcal{P} \sim DP(\alpha, G_0)$$

Distribución posterior:

$$\mathcal{P} | \{X_1, \ldots, X_n\} \sim DP\left(\alpha + n, \frac{\alpha}{\alpha + n} G_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{X_i}\right)$$

[Ferguson, 1973]

Estimación de una distribución

Tomando $\mathcal{A} = \{A_1, \ldots, A_k\}$ una partición cualquiera, se tiene que

$$\mathcal{P}(\mathcal{A})|X_1, \ldots, X_n \sim \mathit{Dirichlet}\left(\alpha.G_0(A_1) + \sum_{i=1}^{n} \mathbb{1}_{\{X_i \in A_1\}}, \ldots, \alpha.G_0(A_k) + \sum_{i=1}^{n} \mathbb{1}_{\{X_i \in A_k\}}\right)$$

Si $\mathcal{A} = \{A_x, A_x^c\}$ con $A_x = (-\infty, x]$, entonces $\mathcal{P}(A_x) = F(x)$:

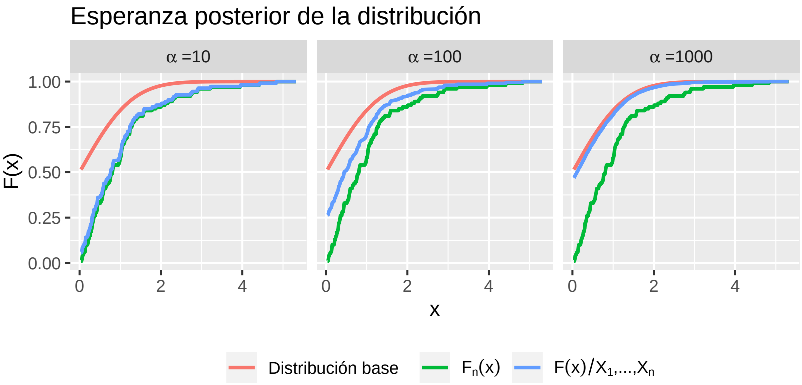
$$F(x)|X_1, \ldots, X_n \sim \mathit{Beta}\Big(\alpha G_o(A_x) + nF_n(x), \alpha(1 - G_0(A_x)) + n(1 - F_n(x))\Big)$$

Estimación de una distribución

- $\mathbb{E}\left(F(x)|X_1,\ldots,X_n\right) = \left(\frac{\alpha}{\alpha+n}\right) G_0\left(A_x\right) + \left(\frac{n}{\alpha+n}\right) F_n(x)$
- $Var(F(x)|X_1,\ldots,X_n) = \frac{(\alpha G_0(A_x)+)(\alpha G_0(A_x^c)+nF_n(x)}{(\alpha+n)^2(\alpha+n+1)}$

Estimación de una distribución

- $\mathbb{E}\left(F(x)|X_1,\ldots,X_n\right) = \left(\frac{\alpha}{\alpha+n}\right) G_0\left(A_x\right) + \left(\frac{n}{\alpha+n}\right) F_n(x)$
- $Var(F(x)|X_1,\ldots,X_n) = \frac{(\alpha G_0(A_x)+)(\alpha G_0(A_x^c)+nF_n(x)}{(\alpha+n)^2(\alpha+n+1)}$



Modelos de mezcla controlados por DP

Si f está dada por un DPMM entonces:

$$
\begin{aligned}
f(x) &= \int f_\theta(x|\theta)d\mathrm{P}(\theta) \\
&= \sum_h \pi_h f_\theta(x|\theta_h)
\end{aligned}
$$

donde $\mathrm{P} = \sum_h \pi_h \delta_{\theta_h}$ es un proceso de Dirichlet.

Modelos de mezcla controlados por DP

Si f está dada por un DPMM entonces:

$$\begin{aligned} f(x) &= \int f_\theta(x|\theta) d\mathrm{P}(\theta) \\ &= \sum_h \pi_h f_\theta(x|\theta_h) \end{aligned}$$

donde $\mathrm{P} = \sum_h \pi_h \delta_{\theta_h}$ es un proceso de Dirichlet.

Esta formulación es equivalente a plantear el siguiente modelo jerárquico:

$$\begin{aligned} &X_i \overset{ind}{\sim} F(\cdot|\theta_i) \\ &\theta_i \overset{ind}{\sim} \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{\theta_h^*} \\ &\left.\begin{aligned} &\theta_h^* \overset{ind}{\sim} G_0 \\ &\pi_h \sim Stick(\alpha) \end{aligned}\right\} \quad \mathcal{P} \sim DP(\alpha, G_0) \end{aligned}$$

[Gelman et al., 2013], [Niemi, 2017]

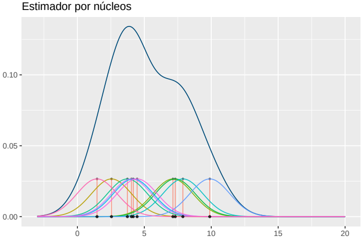
Analogía con los estimadores por núcleo

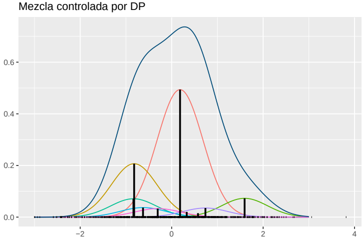
Los estimadores por núcleo

$$f \approx \hat{f} = \sum_{i=1}^{n} \frac{1}{n} K_{h,x_i}(x)$$

En nuestro caso:

$$f \approx \hat{f} = \sum_{h=1}^{\infty} \pi_h K_{\theta_h}(x)$$





Modelos de mezcla DP con distribución normal

$$
\begin{aligned}
&X_i \overset{ind}{\sim} F(\cdot|\theta_i) \\
&\theta_i \overset{ind}{\sim} \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{\theta_h^*} \\
&\left.\begin{aligned} &\theta_h^* \overset{ind}{\sim} G_0 \\ &\pi_h \sim Stick(\alpha) \end{aligned}\right\} \quad \mathcal{P} \sim DP(\alpha, G_0)
\end{aligned}
$$

Modelos de mezcla DP con distribución normal

$$
\begin{aligned}
&X_i \overset{ind}{\sim} F(\cdot|\theta_i) \\
&\theta_i \overset{ind}{\sim} \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{\theta_h^*} \\
&\left.\begin{array}{l} \theta_h^* \overset{ind}{\sim} G_0 \\ \pi_h \sim Stick(\alpha) \end{array}\right\} \quad \mathcal{P} \sim DP(\alpha, G_0)
\end{aligned}
$$

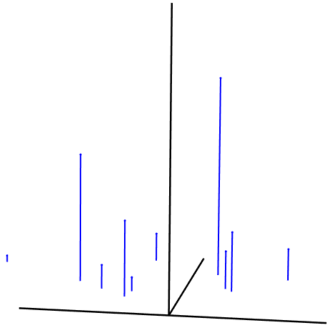
$$
\begin{aligned}
X_i \quad & \overset{ind}{\sim} \mathcal{N}(\mu_i, \sigma_i^2) \\
(\mu_i, \sigma_i^2) \quad & \sim \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{(\mu_h^*, \sigma_h^{2*})} \\
(\mu_h^*, \sigma_h^{2*}) \quad & \sim \underbrace{N(\mu|m, \frac{1}{k}\sigma^2) \times IG(\sigma^2|\nu, \psi)}_{G_0(\mu,\sigma)} \\
\pi_h \quad & \sim Stick(\alpha) \\
m \in \mathbb{R},\ & \alpha, k, \nu, \psi \in \mathbb{R}^+
\end{aligned}
$$

Modelos de mezcla DP con distribución normal

$$
\begin{aligned}
&X_i \overset{ind}{\sim} F(\cdot|\theta_i) \\
&\theta_i \overset{ind}{\sim} \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{\theta_h^*} \\
&\left.\begin{aligned}\theta_h^* &\overset{ind}{\sim} G_0 \\ \pi_h &\sim Stick(\alpha)\end{aligned}\right\} \quad \mathcal{P} \sim DP(\alpha, G_0)
\end{aligned}
$$

$$
\begin{aligned}
X_i \quad & \overset{ind}{\sim} \mathcal{N}(\mu_i, \sigma_i^2) \\
(\mu_i, \sigma_i^2) \quad & \sim \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{(\mu_h^*, \sigma_h^{2*})} \\
(\mu_h^*, \sigma_h^{2*}) \quad & \sim \underbrace{N(\mu|m, \frac{1}{k}\sigma^2) \times IG(\sigma^2|\nu, \psi)}_{G_0(\mu,\sigma)} \\
\pi_h \quad & \sim Stick(\alpha) \\
m \in \mathbb{R},\ \alpha, k, \nu, \psi \in \mathbb{R}^+ &
\end{aligned}
$$

Stick-breaking bidimensional con base G_0



Modelos de mezcla DP con distribución normal

$$
\begin{aligned}
&X_i \overset{ind}{\sim} F(\cdot|\theta_i)\\
&\theta_i \overset{ind}{\sim} \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{\theta_h^*}\\
&\left.\begin{aligned}\theta_h^* &\overset{ind}{\sim} G_0\\ \pi_h &\sim Stick(\alpha)\end{aligned}\right\} \quad \mathcal{P} \sim DP(\alpha, G_0)
\end{aligned}
$$

$$
\begin{aligned}
X_i \quad & \overset{ind}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)\\
(\mu_i, \sigma_i^2) \quad & \sim \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{(\mu_h^*, \sigma_h^{2*})}\\
(\mu_h^*, \sigma_h^{2*}) \quad & \sim \underbrace{N(m, \frac{1}{k}\sigma^2) \times IG(\nu, \psi)}_{G_0(\mu, \sigma)}\\
\pi_h \quad & \sim Stick(\alpha)\\
m \in \mathbb{R},\ \alpha, k, \nu, \psi \in \mathbb{R}^+ &
\end{aligned}
$$

$$X_i \quad \overset{ind}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)$$

Modelos de mezcla DP con distribución normal

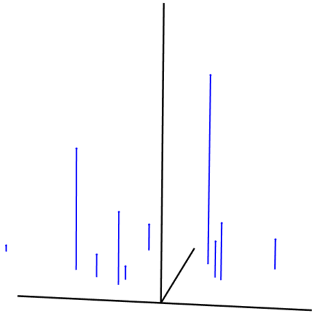
$$
\begin{aligned}
&X_i \overset{ind}{\sim} F(\cdot|\theta_i)\\
&\theta_i \overset{ind}{\sim} \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{\theta_h^*}\\
&\left.\begin{array}{l} \theta_h^* \overset{ind}{\sim} G_0 \\ \pi_h \sim Stick(\alpha) \end{array}\right\} \quad \mathcal{P} \sim DP(\alpha, G_0)
\end{aligned}
$$

$$
\begin{aligned}
X_i \quad & \overset{ind}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)\\
(\mu_i, \sigma_i^2) \quad & \sim \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{(\mu_h^*, \sigma_h^{2*})}\\
(\mu_h^*, \sigma_h^{2*}) \quad & \sim \underbrace{N(m, \frac{1}{k}\sigma^2) \times IG(\nu, \psi)}_{G_0(\mu,\sigma)}\\
\pi_h \quad & \sim Stick(\alpha)\\
m \in \mathbb{R},\ \alpha, k, \nu, \psi \in \mathbb{R}^+ &
\end{aligned}
$$

$$
\begin{aligned}
X_i \quad & \overset{ind}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)\\
(\mu_i, \sigma_i^2) \quad & \sim \mathcal{P} = \textstyle\sum_{h=1}^{H} \pi_h \delta_{(\mu_h^*, \sigma_h^{2*})}
\end{aligned}
$$

Modelos de mezcla DP con distribución normal

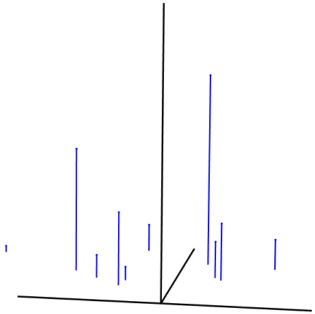
$$
\begin{aligned}
&X_i \overset{ind}{\sim} F(\cdot|\theta_i)\\
&\theta_i \overset{ind}{\sim} \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{\theta_h^*}\\
&\left.\begin{aligned}\theta_h^* &\overset{ind}{\sim} G_0\\ \pi_h &\sim Stick(\alpha)\end{aligned}\right\} \quad \mathcal{P} \sim DP(\alpha, G_0)
\end{aligned}
$$



$$
\begin{aligned}
X_i \quad &\overset{ind}{\sim} \mathcal{N}(\mu_i, \sigma_i^2)\\
(\mu_i, \sigma_i^2) \quad &\sim \mathcal{P} = \underbrace{\sum_{h=1}^{H} \pi_h \delta_{(\mu_h^*, \sigma_h^{2*})}}_{\text{Distribución discreta}}
\end{aligned}
$$

Modelos de mezcla DP con distribución normal

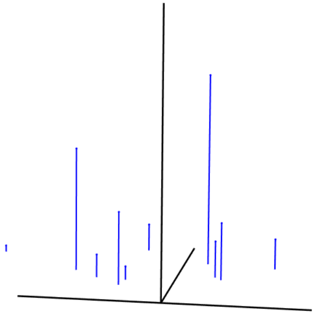
$$X_i \overset{ind}{\sim} F(\cdot|\theta_i)$$
$$\theta_i \overset{ind}{\sim} \mathcal{P} = \sum_{h=1}^{\infty} \pi_h \delta_{\theta_h^*}$$
$$\left.\begin{array}{l} \theta_h^* \overset{ind}{\sim} G_0 \\ \pi_h \sim Stick(\alpha) \end{array}\right\} \quad \mathcal{P} \sim DP(\alpha, G_0)$$



$$\begin{aligned} X_i &\overset{ind}{\sim} \mathcal{N}(\mu_i, \sigma_i^2) \\ (\mu_i, \sigma_i^2) &\sim \mathcal{P} = \sum_{h=1}^{H} \pi_h \delta_{(\mu_h^*, \sigma_h^{2*})} \\ \xi_i &\sim Cat(\{1, \ldots, H\}, \boldsymbol{\pi}) \end{aligned}$$

Modelos de mezcla DP con distribución normal

$$\begin{array}{l} X_i \overset{ind}{\sim} F(\cdot|\theta_i) \\ \theta_i \overset{ind}{\sim} \mathcal{P} = \sum_{h=1}^{\infty} \pi_h \delta_{\theta_h^*} \\ \left.\begin{array}{l} \theta_h^* \overset{ind}{\sim} G_0 \\ \pi_h \sim Stick(\alpha) \end{array}\right\} \quad \mathcal{P} \sim DP(\alpha, G_0) \end{array}$$



$$\begin{aligned} X_i \quad & \overset{ind}{\sim} \mathcal{N}(\mu_i, \sigma_i^2) \\ (\mu_i, \sigma_i^2) \quad & \sim \mathcal{P} = \sum_{h=1}^{H} \pi_h \delta_{(\mu_h^*, \sigma_h^{2*})} \\ \xi_i \quad & \sim Cat(\{1, \ldots, H\}, \boldsymbol{\pi}) \\ (\mu_i, \sigma_i^2)|\{\xi_i = h\} \quad & \overset{cs}{=} (\mu_h^*, \sigma_h^{2*}) \end{aligned}$$

El modelo final

$$
\begin{aligned}
X_i &\overset{ind}{\sim} F(\cdot|\theta_i) \\
\theta_i &\overset{ind}{\sim} \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{\theta_h^*} \\
&\left.\begin{aligned} \theta_h^* &\overset{ind}{\sim} G_0 \\ \pi_h &\sim Stick(\alpha) \end{aligned}\right\} \quad \mathcal{P} \sim DP(\alpha, G_0)
\end{aligned}
$$

$$
\begin{aligned}
X_i \quad & \overset{ind}{\sim} \mathcal{N}(\mu_i, \sigma_i^2) \\
(\mu_i, \sigma_i^2) \quad & \sim \mathcal{P} = \textstyle\sum_{h=1}^{\infty} \pi_h \delta_{(\mu_h^*, \sigma_h^{2*})} \\
(\mu_h^*, \sigma_h^{2*}) \quad & \sim \underbrace{N(\mu|m, \frac{1}{k}\sigma^2) \times IG(\sigma^2|\nu, \psi)}_{G_0(\mu, \sigma)} \\
\pi_h \quad & \sim Stick(\alpha) \\
m \in \mathbb{R}, \ & \alpha, k, \nu, \psi \in \mathbb{R}^+
\end{aligned}
$$

$$
\begin{aligned}
X_i &\overset{ind}{\sim} \sum_{h=1}^{H} \pi_h N(\mu_h^*, \sigma_h^{2*}) \\
(\mu_h^*, \sigma_h^{2*}) &\sim N(m, \frac{1}{k}\sigma^2) \times IG(\nu, \psi) \\
\pi_h &\sim Stick(\alpha) \\
m \in \mathbb{R}, \ & \alpha, k, \nu, \psi \in \mathbb{R}^+
\end{aligned}
$$

Obtención de estimadores

$$X_i \overset{ind}{\sim} \sum_{h=1}^{H} \pi_h N(\mu_h^*, \sigma_h^{2*})$$

La densidad queda determinada por los parámetros $\{\pi_h, \mu_h^*, \sigma_h^{2*}\}_{h=1,\ldots,H}$

Obtención de estimadores

$$X_i \overset{ind}{\sim} \sum_{h=1}^{H} \pi_h N(\mu_h^*, \sigma_h^{2*})$$

La densidad queda determinada por los parámetros $\{\pi_h, \mu_h^*, \sigma_h^{2*}\}_{h=1,\ldots,H}$

El algoritmo de Gibbs devuelve:

$$\begin{array}{c|c|c|c|c}
\overbrace{\hphantom{\mu_{1,1}}}^{\text{Iter 1}} & \overbrace{\hphantom{\mu_{1,2}}}^{\text{Iter 2}} & \overbrace{\hphantom{\mu_{1,3}}}^{\text{Iter 3}} & & \overbrace{\hphantom{\mu_{1,m}}}^{\text{Iter m}} \\
\pi_{1,1} & \pi_{1,2} & \pi_{1,3} & \cdots & \pi_{1,m} \\
\mu_{1,1} & \mu_{1,2} & \mu_{1,3} & \cdots & \mu_{1,m} \\
\sigma_{1,1} & \sigma_{1,2} & \sigma_{1,3} & \cdots & \sigma_{1,m} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\pi_{H,1} & \pi_{H,2} & \pi_{H,3} & \cdots & \pi_{H,m} \\
\mu_{H,1} & \mu_{H,2} & \mu_{H,3} & \cdots & \mu_{H,m} \\
\sigma_{H,1} & \sigma_{H,2} & \sigma_{H,3} & \cdots & \sigma_{H,m}
\end{array}$$

Obtención de estimadores

$$X_i \overset{ind}{\sim} \sum_{h=1}^{H} \pi_h N(\mu_h^*, \sigma_h^{2*})$$

La densidad queda determinada por los parámetros $\{\pi_h, \mu_h^*, \sigma_h^{2*}\}_{h=1,\ldots,H}$

El algoritmo de Gibbs devuelve:

Iter 1	Iter 2	Iter 3		Iter m
$\pi_{1,1}$	$\pi_{1,2}$	$\pi_{1,3}$	$\cdots$	$\pi_{1,m}$
$\mu_{1,1}$	$\mu_{1,2}$	$\mu_{1,3}$	$\cdots$	$\mu_{1,m}$
$\sigma_{1,1}$	$\sigma_{1,2}$	$\sigma_{1,3}$	$\cdots$	$\sigma_{1,m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\pi_{H,1}$	$\pi_{H,2}$	$\pi_{H,3}$	$\cdots$	$\pi_{H,m}$
$\mu_{H,1}$	$\mu_{H,2}$	$\mu_{H,3}$	$\cdots$	$\mu_{H,m}$
$\sigma_{H,1}$	$\sigma_{H,2}$	$\sigma_{H,3}$	$\cdots$	$\sigma_{H,m}$

En cada iteración:

$$\hat{f}(x)^{(k)} = \sum_{h=1}^{H} \pi_h{}^{(k)} \phi(x; \mu_h{}^{(k)}, \sigma_h{}^{(k)})$$

Obtención de estimadores

$$X_i \overset{ind}{\sim} \sum_{h=1}^{H} \pi_h N(\mu_h^*, \sigma_h^{2*})$$

La densidad queda determinada por los parámetros $\{\pi_h, \mu_h^*, \sigma_h^{2*}\}_{h=1,\ldots,H}$

El algoritmo de Gibbs devuelve:

Iter 1	Iter 2	Iter 3		Iter m
$\pi_{1,1}$	$\pi_{1,2}$	$\pi_{1,3}$	$\cdots$	$\pi_{1,m}$
$\mu_{1,1}$	$\mu_{1,2}$	$\mu_{1,3}$	$\cdots$	$\mu_{1,m}$
$\sigma_{1,1}$	$\sigma_{1,2}$	$\sigma_{1,3}$	$\cdots$	$\sigma_{1,m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\pi_{H,1}$	$\pi_{H,2}$	$\pi_{H,3}$	$\cdots$	$\pi_{H,m}$
$\mu_{H,1}$	$\mu_{H,2}$	$\mu_{H,3}$	$\cdots$	$\mu_{H,m}$
$\sigma_{H,1}$	$\sigma_{H,2}$	$\sigma_{H,3}$	$\cdots$	$\sigma_{H,m}$

En cada iteración:

$$\hat{f}(x)^{(k)} = \sum_{h=1}^{H} {\pi_h}^{(k)} \phi(x; {\mu_h}^{(k)}, {\sigma_h}^{(k)})$$

Un estimador puntual:

$$\hat{f}(x) = \frac{1}{m} \sum_{k=1}^{m} \hat{f}(x)^{(k)}$$

Obtención de estimadores

$$X_i \overset{ind}{\sim} \sum_{h=1}^{H} \pi_h N(\mu_h^*, \sigma_h^{2*})$$

La densidad queda determinada por los parámetros $\{\pi_h, \mu_h^*, \sigma_h^{2*}\}_{h=1,\ldots,H}$

El algoritmo de Gibbs devuelve:

Iter 1	Iter 2	Iter 3		Iter m
$\pi_{1,1}$	$\pi_{1,2}$	$\pi_{1,3}$	$\cdots$	$\pi_{1,m}$
$\mu_{1,1}$	$\mu_{1,2}$	$\mu_{1,3}$	$\cdots$	$\mu_{1,m}$
$\sigma_{1,1}$	$\sigma_{1,2}$	$\sigma_{1,3}$	$\cdots$	$\sigma_{1,m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\pi_{H,1}$	$\pi_{H,2}$	$\pi_{H,3}$	$\cdots$	$\pi_{H,m}$
$\mu_{H,1}$	$\mu_{H,2}$	$\mu_{H,3}$	$\cdots$	$\mu_{H,m}$
$\sigma_{H,1}$	$\sigma_{H,2}$	$\sigma_{H,3}$	$\cdots$	$\sigma_{H,m}$

En cada iteración:

$$\hat{f}(x)^{(k)} = \sum_{h=1}^{H} \pi_h{}^{(k)} \phi(x; \mu_h{}^{(k)}, \sigma_h{}^{(k)})$$

Un estimador puntual:

$$\hat{f}(x) = \frac{1}{m} \sum_{k=1}^{m} \hat{f}(x)^{(k)}$$

Si en lugar de promediar, se toman percentiles, se obtiene un intervalo de credibilidad.

Estudio de simulación

Caso 1: χ^2_{10}: Distribución χ^2 con 10 grados de libertad.

Caso 2: Claw: $f(x) = \frac{1}{2}\phi(x; 0, 1) + \frac{1}{10}\sum_{j=0}^{4}\phi(x; \frac{j}{2} - 1, \frac{1}{10})$

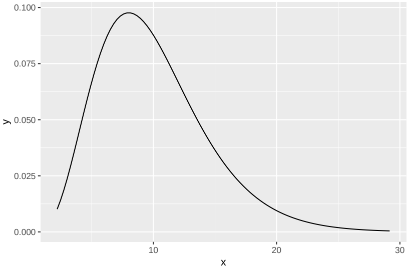
Caso 3: Mix$_1$: $f(x) = \frac{1}{3}\phi(x; -4, 0.1) + \frac{1}{3}\phi(x; -2, 0.5) + \frac{1}{3}\phi(x; 2, 1)$

Caso 4: Mix$_2$: $f(x) = \frac{1}{2}\mathbb{1}_{\{[-2,-1]\}}(x) + \frac{1}{2}\mathbb{1}_{\{[1,2]\}}(x)$

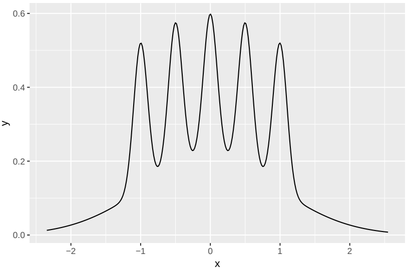
[Bourel & Cugliari, 2018]

$$\mathrm{MISE}(\hat{f}) = \mathbb{E}\left[\int_{\mathbb{R}}(\hat{f}(x) - f(x))^2 dx\right]$$

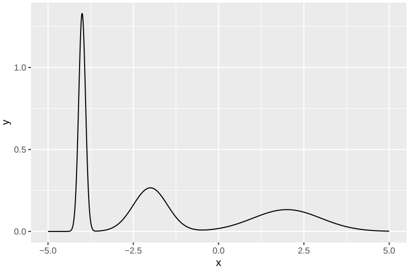
Distribuciones a simular



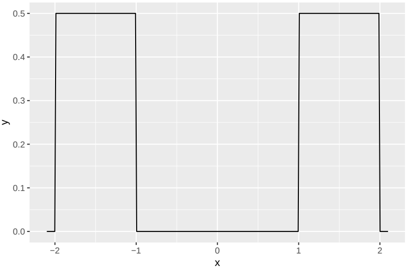
(a) χ^2_{10}



(b) Claw

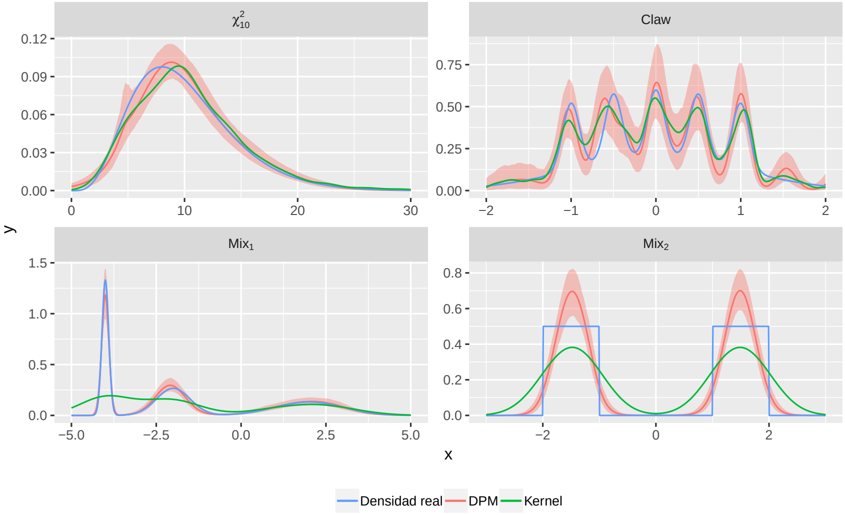


(c) Mix_1



(d) Mix_2

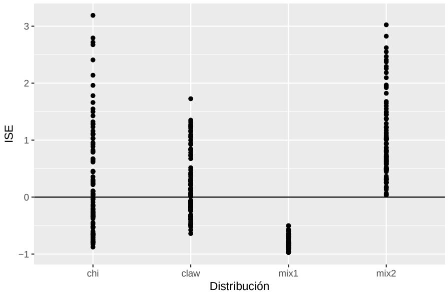
Estimaciones obtenidas



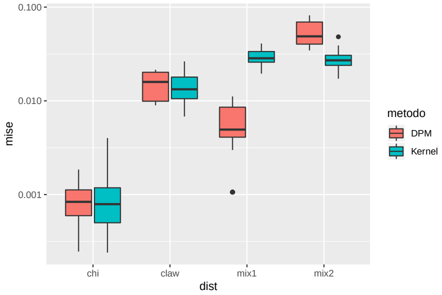
MISE

	DPM	Kernel
χ^2_{10}	**8.53E-04**	9.42E-04
Claw	1.52E-02	**1.44E-02**
Mix_1	**5.71E-03**	2.94E-02
Mix_2	5.32E-02	**2.75E-02**

Cuadro: MISE en cada uno de los casos



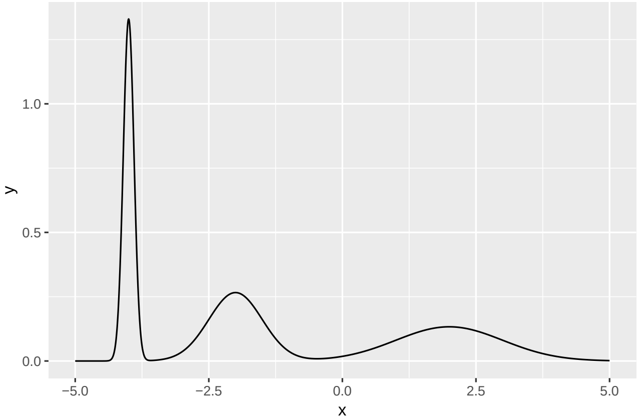
(a) Diferencia relativa entre el error de DPM y Kernel (DPM-Ker)/Ker



(b) Gráficos de caja de los valores del error cuadrático integrado

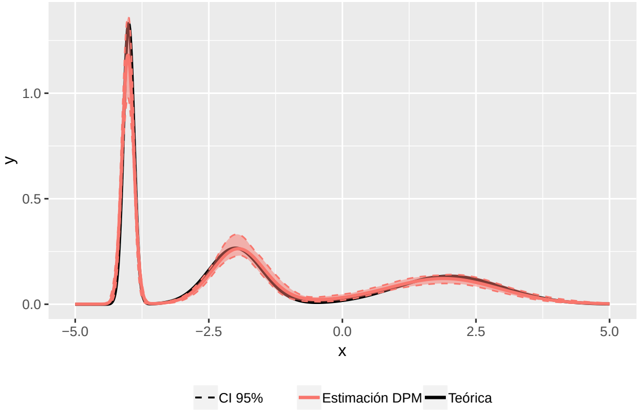
Un ejemplo

$$f(x) = \frac{1}{3}\phi(x; -4, 0.1) + \frac{1}{3}\phi(x; -2, 0.5) + \frac{1}{3}\phi(x; 2, 1)$$



Un ejemplo

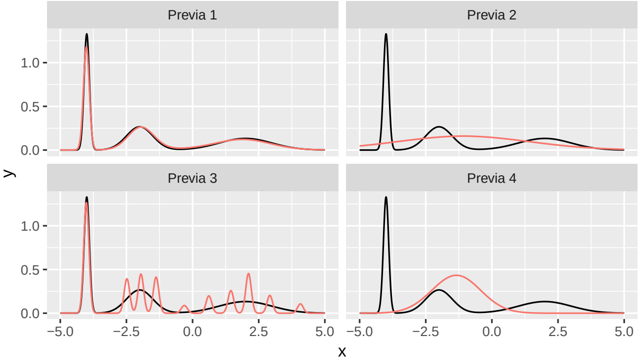
$$f(x) = \frac{1}{3}\phi(x; -4, 0.1) + \frac{1}{3}\phi(x; -2, 0.5) + \frac{1}{3}\phi(x; 2, 1)$$



Parámetros de la distinta distribución base

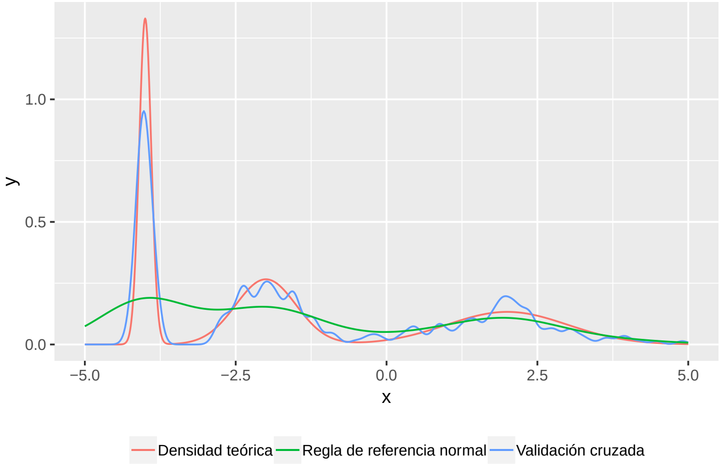
Caso 1	$\alpha = 1$	$m = -1$	$k = 1$	$\nu = 0.1$	$\psi = 0.1$
Caso 2	$\alpha = 1$	$m = -1$	$k = 0.01$	$\nu = 0.1$	$\psi = 0.1$
Caso 3	$\alpha = 1$	$m = -1$	$k = 20$	$\nu = 1000$	$\psi = 10$
Caso 4	$\alpha = 1$	$m = -1$	$k = 0.1$	$\nu = 1000$	$\psi = 10$

Cuadro: Parámetros de la distribución base usados

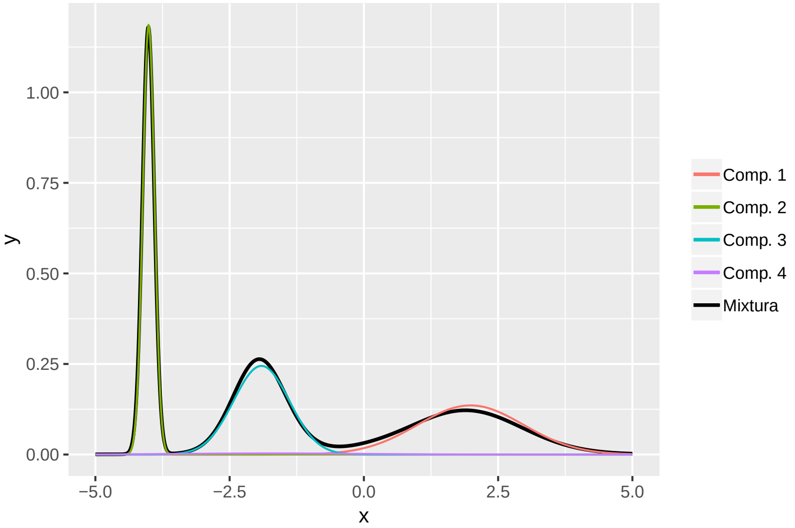


Estimador kernel

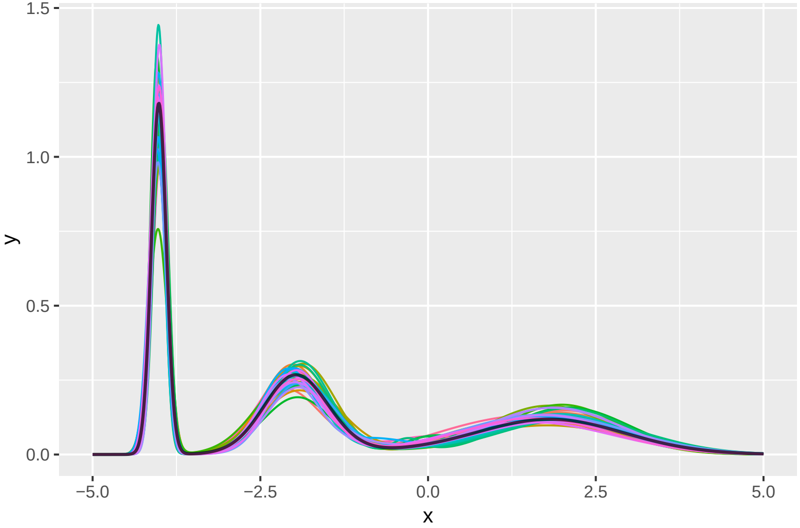
El método de estimación por kernel al tener ancho de banda fijo no produce un buen resultado.



Componentes de la mezcla



La estimación en distintas iteraciones



Resumen y conclusiones

- Se quiso estimar la distribución de una v.a. con recorrido $\{1, 2, \ldots, k\}$
  $\Rightarrow$ Modelo Dirichlet-Multinomial

Resumen y conclusiones

- Se quiso estimar la distribución de una v.a. con recorrido $\{1, 2, \ldots, k\}$
  $\Rightarrow$ Modelo Dirichlet-Multinomial

- Extensión a variables con recorrido infinito.
  $\Rightarrow$ Los datos se distribuyen según una medida de probabilidad aleatoria, proceso de Dirichlet como previa y posterior. Modelo conjugado.

Resumen y conclusiones

- Se quiso estimar la distribución de una v.a. con recorrido $\{1, 2, \ldots, k\}$
  $\Rightarrow$ Modelo Dirichlet-Multinomial

- Extensión a variables con recorrido infinito.
  $\Rightarrow$ Los datos se distribuyen según una medida de probabilidad aleatoria, proceso de Dirichlet como previa y posterior. Modelo conjugado.
  Limitante: Estimación discreta

- Variables aleatorias con densidad
  $\Rightarrow$ Modelos de mezcla controlados por el proceso de Dirichlet.

Resumen y conclusiones

- Se quiso estimar la distribución de una v.a. con recorrido $\{1, 2, \ldots, k\}$
  $\Rightarrow$ Modelo Dirichlet-Multinomial

- Extensión a variables con recorrido infinito.
  $\Rightarrow$ Los datos se distribuyen según una medida de probabilidad aleatoria, proceso de Dirichlet como previa y posterior. Modelo conjugado.
  Limitante: Estimación discreta

- Variables aleatorias con densidad
  $\Rightarrow$ Modelos de mezcla controlados por el proceso de Dirichlet.

- Comparado con el estimador por núcleos:
  - Tiene un desempeño aceptable (MISE).
  - Permite obtener bandas de credibilidad.
  - Es más costoso en tiempo computacional.

Trabajo a futuro

- Extender la estimación al caso multidimensional. Para ello considerar G_0 como Normal multivariada - Inversa Wishart.
- Complejizar el estudio de simulación, variando el tamaño de muestra n.
- Investigar otras posibilidades para $F(\cdot|\theta_i)$ que no sean la distribución normal.
- Asignar una distribución previa a los parámetros de la medida base. (Implementación en **DPpackage**).
- En lugar de promediar iteraciones, tomar la curva "más profunda" en el sentido de alguna función de profundidad.
- Uso de esta técnica en clustering a través de las variables latentes.

Referencias

Bourel, M. and Cugliari, J. (2018).
Bagging of density estimators.
arXiv preprint arXiv:1808.03447.

Ferguson, T. S. (1973).
A bayesian analysis of some nonparametric problems.
The annals of statistics, pages 209–230.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013).
Bayesian data analysis.
Chapman and Hall/CRC.

Niemi, J. (2017).
Bayesian nonparametrics.
http://www.jarad.me/courses/stat615/slides/Nonparametrics/nonparametrics.pdf.

Sethuraman, J. (1994).
A constructive definition of dirichlet priors.
Statistica sinica, pages 639–650.

Agradecimientos

¡Muchas gracias!

Estimación de densidades mediante modelos de mezcla controlados por el proceso de Dirichlet

Estudiantes:

Manuel Hernández Banadik
Mario Sierra Golombievski

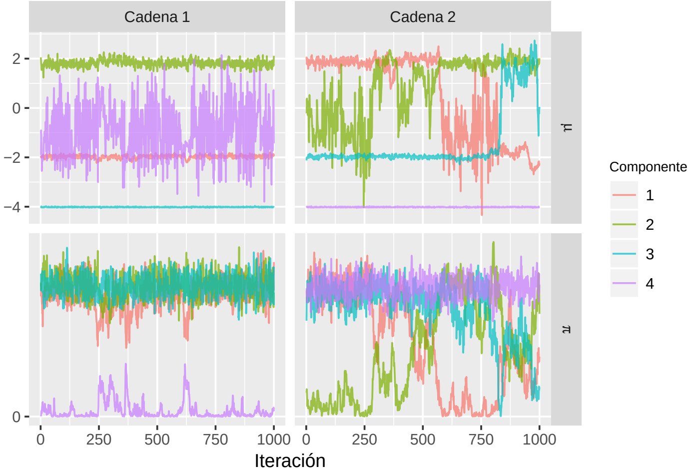
Tutor:

Ignacio Alvarez-Castro

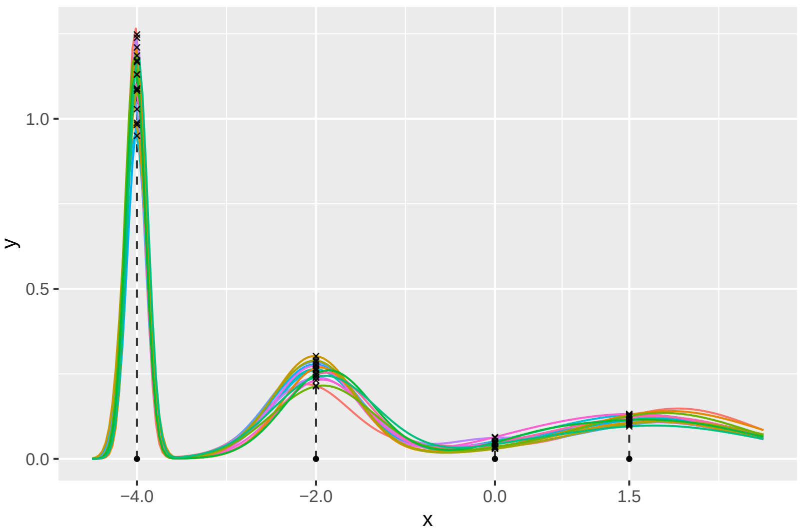
2 de julio de 2019

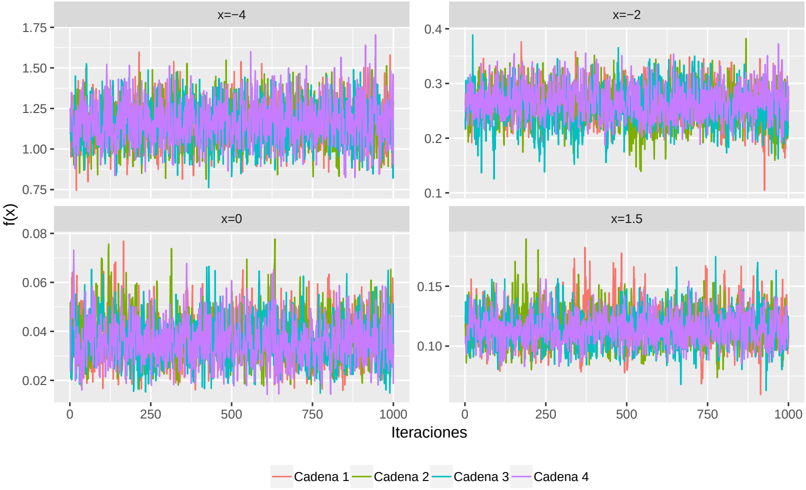
Monitoreo de convergencia

Por cuestiones de identificabilidad el monitoreo no se hace a nivel de cada parámetro. Los parámetros pueden *intercambiar sus índices* a lo largo de las iteraciones sin que esto implique un problema de convergencia:



Para monitorear la convergencia podemos monitorear la convergencia puntual de cada cadena de estimadores evaluado en una grilla de puntos, en este caso $\{x = -4, x = 2, x = 0, x = 1.5\}$




x=−4
x=−2
x=0
x=1.5
f(x)
Iteraciones
1.75
1.50
1.25
1.00
0.75
0.4
0.3
0.2
0.1
0.08
0.06
0.04
0.02
0.15
0.10
0
250
500
750
1000
Cadena 1
Cadena 2
Cadena 3
Cadena 4