Reducción Suficientes de Dimensiones en Regresión con Predictores Mixtos

Pamela Llop

Trabajo en colaboración con Liliana Forzani, Rodrigo García Arancibia y Diego Tomassi

UMA 2016 - Bahía Blanca

FIQ UNL

Problema

Asignar ayuda económica a hogares o individuos que viven en situación de **pobreza**

Políticas o Programas Focalizados
(Ejemplos reales: CAS in Chile, Sisben en Colombia, SISFOH en Perú, Tekoporá en Paraguay, SIERP en Honduras, PANES en Uruguay, entre otros)

Problema

Asignar ayuda económica a hogares o individuos que viven en situación de **pobreza**

Un hogar j es pobre si $Y_j \leq LP$ donde LP es el ingreso que determina la **línea de pobreza** (valor monetario de la CBT)

Políticas o Programas Focalizados
(Ejemplos reales: CAS in Chile, Sisben en Colombia, SISFOH en Perú, Tekoporá en Paraguay, SIERP en Honduras, PANES en Uruguay, entre otros)

Problema

Asignar ayuda económica a hogares o individuos que viven en situación de **pobreza**

Un hogar j es pobre si $Y_j \leq LP$ donde LP es el ingreso que determina la **línea de pobreza** (valor monetario de la CBT)

$Y_j \leq LP \rightsquigarrow$ asigna la ayuda

Políticas o Programas Focalizados

(Ejemplos reales: CAS in Chile, Sisben en Colombia, SISFOH en Perú, Tekoporá en Paraguay, SIERP en Honduras, PANES en Uruguay, entre otros)

Problema

Asignar ayuda económica a hogares o individuos que viven en situación de **pobreza**

Un hogar j es pobre si $Y_j \leq LP$ donde LP es el ingreso que determina la **línea de pobreza** (valor monetario de la CBT)

$Y_j \leq LP \rightsquigarrow$ asigna la ayuda

Y_j es una variable muy difícil de medir:

- Incentivos a revelar el valor verdadero.
- Economía informal (ej. trueques, auto-provisiones).
- Estacionalidad (ej. changas, empleos rurales).

Situación real: construir un índice

- **Indice de estatus socioeconómico**

$$I_j = \alpha_1 X_{1j} + \cdots + \alpha_p X_{pj} = \boldsymbol{\alpha}^T \mathbf{X}_j$$

Situación real: construir un índice

- **Indice de estatus socieconómico**

$$I_j = \alpha_1 X_{1j} + \cdots + \alpha_p X_{pj} = \boldsymbol{\alpha}^T \mathbf{X}_j$$

- Otras variables $\mathbf{X} = (X_1, \ldots, X_p)$ que son más fáciles de medir:
  - Vivienda (materiales del techo, materiales del suelo, forma de acceso al agua potable, etc.).
  - Activos físicos (¿tiene: radio?, TV?, internet?, moto?, auto?)
  - Otras socio-demográficas (cant. de miembros, escolaridad, situación ocupacional, etc.)

Situación real: construir un índice

- **Indice de estatus socieconómico**

$$I_j = \alpha_1 X_{1j} + \cdots + \alpha_p X_{pj} = \boldsymbol{\alpha}^T \mathbf{X}_j$$

- Otras variables $\mathbf{X} = (X_1, \ldots, X_p)$ que son más fáciles de medir:
  - Vivienda (materiales del techo, materiales del suelo, forma de acceso al agua potable, etc.).
  - Activos físicos (¿tiene: radio?, TV?, internet?, moto?, auto?)
  - Otras socio-demográficas (cant. de miembros, escolaridad, situación ocupacional, etc.)

$\boldsymbol{\alpha_1, \alpha_2, \ldots ???}$

Cómo construir el índice

Reducción Suficiente de Dimensiones (RSD)

↓

Reducir la dimensión de un vector de predictores sin perder información sobre la respuesta

Cómo construir el índice

Reducción Suficiente de Dimensiones (RSD)

⬇

Reducir la dimensión de un vector de predictores sin perder información sobre la respuesta

⬆

Gran ventaja sobre PCA!!!

Reducción suficiente de dimensiones

Sea $\mathbf{X} \in \mathbb{R}^p, Y \in \mathbb{R}$, una reducción $R(\mathbf{X}) \in \mathbb{R}$ es suficiente para la regresión de $Y|\mathbf{X}$

⇕

$$Y|\mathbf{X} \sim Y|R(\mathbf{X})$$

Reducción suficiente de dimensiones

Sea $\mathbf{X} \in \mathbb{R}^p, Y \in \mathbb{R}$, una reducción $R(\mathbf{X}) \in \mathbb{R}$ es suficiente para la regresión de $Y|\mathbf{X}$

$$\Updownarrow$$

$$Y|\mathbf{X} \sim Y|R(\mathbf{X})$$

$$\Updownarrow$$

$$\mathbf{X}|(R(\mathbf{X}), Y) \sim \mathbf{X}|R(\mathbf{X})$$

Reducción suficiente de dimensiones

Sea $\mathbf{X} \in \mathbb{R}^p, Y \in \mathbb{R}$, una reducción $R(\mathbf{X}) \in \mathbb{R}$ es suficiente para la regresión de $Y|\mathbf{X}$

$$\Updownarrow$$

$$Y|\mathbf{X} \sim Y|R(\mathbf{X})$$

$$\Updownarrow$$

$$\mathbf{X}|(R(\mathbf{X}), Y) \sim \mathbf{X}|R(\mathbf{X})$$

RSD $\rightsquigarrow$ basada en la distribución de $\mathbf{X}|Y$ sin necesidad de asumir alguna distribución para $Y|\mathbf{X}$.

Reducción suficiente de dimensiones

- $\mathbf{X}|Y \sim N(\boldsymbol{\mu} + \boldsymbol{\Delta}\boldsymbol{\alpha}\boldsymbol{\xi}(f_Y - \bar{f}_Y), \boldsymbol{\Delta})$
  - $R(\mathbf{X}) = \boldsymbol{\alpha}^T\mathbf{X}$, con $\boldsymbol{\alpha} = \mathrm{span}\{\boldsymbol{\mu}_Y - \boldsymbol{\mu}, Y \in S_Y\}$

    (*Cook*, 2007; *Cook & Forzani*, 2008)

Reducción suficiente de dimensiones

- $\mathbf{X}|Y \sim N(\boldsymbol{\mu} + \boldsymbol{\Delta}\boldsymbol{\alpha}\boldsymbol{\xi}(f_Y - \bar{f}_Y), \boldsymbol{\Delta})$
  - $R(\mathbf{X}) = \boldsymbol{\alpha}^T\mathbf{X}$, con $\boldsymbol{\alpha} = \operatorname{span}\{\boldsymbol{\mu}_Y - \boldsymbol{\mu}, Y \in S_Y\}$

    (*Cook*, 2007; *Cook & Forzani*, 2008)

- $f(\mathbf{X}|Y) = h(\mathbf{X})\exp\{T(\mathbf{X})^T\boldsymbol{\eta}_Y - \psi(\boldsymbol{\eta}_Y)\}$
  - $\mathbf{R}(\mathbf{X}) = \mathbf{a}^T(\mathbf{T}(\mathbf{X}) - \mathrm{E}(\mathbf{T}(\mathbf{X})))$, $\mathbf{a} = \operatorname{span}\{\boldsymbol{\eta}_Y - \boldsymbol{\eta}, Y \in S_Y\}$

    (*Cook & Li*, 2009; *Bura, Duarte & Forzani*, 2015)

Volviendo al problema de interés...

- *pisos* = 1 (tierra/ladrillos sueltos), 2 (cemento o ladrillo fijo), 3 (mosaicos o baldosas), 4(madera/cerámica y alfombra)
- *techo* = 1 (caña o paja), 2 (chapa de cartón), 3 (chapa de fibrocemento o de metal)
- *agua* = 1 (perforación con bomba manual), 2 (perforación con bomba a motor), 3 (red pública)
- *baño* = 1 (letrina), 2 (inodoro sin botón o cadena -balde), 3 (inodoro botón/cadena/mochila)
- *escolaridad* = 1 (sin instrucción), 2 (primaria incompleta), 3 (primaria completa), 4 (secundaria incompleta)

Volviendo al problema de interés...

- *pisos* = 1 (tierra/ladrillos sueltos), 2 (cemento o ladrillo fijo), 3 (mosaicos o baldosas), 4(madera/cerámica y alfombra)
- *techo* = 1 (caña o paja), 2 (chapa de cartón), 3 (chapa de fibrocemento o de metal)
- *agua* = 1 (perforación con bomba manual), 2 (perforación con bomba a motor), 3 (red pública)
- *baño* = 1 (letrina), 2 (inodoro sin botón o cadena -balde), 3 (inodoro botón/cadena/mochila)
- *escolaridad* = 1 (sin instrucción), 2 (primaria incompleta), 3 (primaria completa), 4 (secundaria incompleta)

Variables Categóricas Ordinales

RSD para $\mathbf{X}$ ordinales

RSD para predictores ordinales

RSD para $\mathbf{X}$ ordinales

Forzani, García Arancibia, Llop & Tomassi (2016)

(PFCord)

Pero además podemos tener...

Variables Ordinales

X

Pero además podemos tener...

- *Sexo jefe/jefa de hogar* = 0 (mujer), 1 (hombre)
- *TV-PC-laptop-auto-moto-celular*= 0 (no tiene), 1 (si tiene)
- *Situación laboral* = 0 (ocupado), 1 (desocupado)

Variables Ordinales + Dicotómicas

$\mathbf{X}$ $\mathbf{H}$

Pero además podemos tener...

- *Sexo jefe/jefa de hogar* = 0 (mujer), 1 (hombre)
- *TV-PC-laptop-auto-moto-celular*= 0 (no tiene), 1 (si tiene)
- *Situación laboral* = 0 (ocupado), 1 (desocupado)
- *Cantidad de personas por metro cuadrado de la vivienda*
- *Cantidad de horas trabajadas por semana*
- *Edad*

Variables Ordinales + Dicotómicas + Continuas

$\mathbf{X}$ $\mathbf{H}$ $\mathbf{W}$

RSD para $\mathbf{X}, \mathbf{W}, \mathbf{H}$ <u>mixtos</u>

RSD para predictores mixtos

RSD para $\mathbf{X}, \mathbf{W}, \mathbf{H}$ mixtos

⬇

Modelo para $\mathbf{X}, \mathbf{W}, \mathbf{H}|Y$

RSD para predictores mixtos

RSD para $\mathbf{X}, \mathbf{W}, \mathbf{H}$ <u>mixtos</u>

$\Downarrow$

Modelo para $\mathbf{X}, \mathbf{W}, \mathbf{H}|Y$

$\Downarrow$

$$f(\mathbf{X}, \mathbf{W}, \mathbf{H}|Y) = f(\mathbf{X}, \mathbf{W}|\mathbf{H}, Y)f(\mathbf{H}|Y)$$

Modelo Bernoulli multivariado

- $\mathbf{H}|Y \sim \text{Bern}(\boldsymbol{\Gamma}^Y)$:

$$P(\mathbf{H}|Y) = P(H_1, \ldots, H_q|Y = y)$$
$$= \frac{1}{G(\boldsymbol{\Gamma}^Y)} \exp\left\{\sum_{j=1}^{q} \gamma_{jj}^Y H_j + \sum_{1 \leq j < j' \leq q} \gamma_{jj'}^Y H_j H_{j'}\right\}$$

- Ising con $[\boldsymbol{\Gamma}^Y]_{ij} = \gamma_{ij}^Y$
  - $\gamma_{jj}^Y = \log\left(\frac{\Pr(H_j=1|\mathbf{H}_{-j}=\mathbf{0},y)}{1-\Pr(H_j=1|\mathbf{H}_{-j}=\mathbf{0},y)}\right)$
  - $\gamma_{jj'}^Y = \log \frac{\Pr(H_j=1,H_{j'}=1|\mathbf{H}_{-j,-j'}=\mathbf{0},y)\Pr(H_j=0,H_{j'}=0|\mathbf{H}_{-j,-j'}=\mathbf{0},y)}{\Pr(H_j=1,H_{j'}=0|\mathbf{H}_{-j,-j'}=\mathbf{0},y)\Pr(H_j=0,H_{j'}=1|\mathbf{H}_{-j,-j'}=\mathbf{0},y)}$

Modelo Bernoulli multivariado

- $\mathbf{H}|Y \sim \text{Bern}(\mathbf{\Gamma}^Y)$:

$$
\begin{aligned}
P(\mathbf{H}|Y) &= P(H_1, \ldots, H_q|Y = y) \\
&= \frac{1}{G(\mathbf{\Gamma}^Y)} \exp\left\{ \sum_{j=1}^{q} \gamma_{jj}^Y H_j + \sum_{1 \le j < j' \le q} \gamma_{jj'}^Y H_j H_{j'} \right\} \\
&= \frac{1}{G(\mathbf{\Gamma}^Y)} \exp\left\{ \text{vech}^T(\mathbf{H}\mathbf{H}^T)\, \text{vech}(\mathbf{\Gamma}^Y) \right\}
\end{aligned}
$$

- Ising con $[\mathbf{\Gamma}^Y]_{ij} = \gamma_{ij}^Y$
  - $\gamma_{jj}^Y = \log\left(\frac{\Pr(H_j=1|\mathbf{H}_{-j}=\mathbf{0},y)}{1-\Pr(H_j=1|\mathbf{H}_{-j}=\mathbf{0},y)} \right)$
  - $\gamma_{jj'}^Y = \log \frac{\Pr(H_j=1,H_{j'}=1|\mathbf{H}_{-j,-j'}=\mathbf{0},y)\,\Pr(H_j=0,H_{j'}=0|\mathbf{H}_{-j,-j'}=\mathbf{0},y)}{\Pr(H_j=1,H_{j'}=0|\mathbf{H}_{-j,-j'}=\mathbf{0},y)\,\Pr(H_j=0,H_{j'}=1|\mathbf{H}_{-j,-j'}=\mathbf{0},y)}$

Modelo Bernoulli multivariado

- $\mathbf{H}|Y \sim$ Bern$(\boldsymbol{\Gamma}^Y)$:

$$
\begin{aligned}
P(\mathbf{H}|Y) &= P(H_1, \ldots, H_q | Y = y) \\
&= \frac{1}{G(\boldsymbol{\Gamma}^Y)} \exp\left\{ \sum_{j=1}^{q} \gamma_{jj}^Y H_j + \sum_{1 \le j < j' \le q} \gamma_{jj'}^Y H_j H_{j'} \right\} \\
&= \frac{1}{G(\boldsymbol{\Gamma}^Y)} \exp\left\{ \mathrm{vech}^T(\mathbf{H}\mathbf{H}^T)\, \mathrm{vech}(\boldsymbol{\Gamma}^Y) \right\} \\
&= \frac{1}{G(\boldsymbol{\Gamma}^Y)} \exp\left\{ \mathrm{vech}^T(\mathbf{H}\mathbf{H}^T) [\boldsymbol{\tau}_0 + \boldsymbol{\tau}\boldsymbol{\kappa}(f_Y - \bar{f}_Y)] \right\}
\end{aligned}
$$

- $\mathrm{vech}(\boldsymbol{\Gamma}^Y) = \boldsymbol{\tau}_0 + \boldsymbol{\tau}\boldsymbol{\kappa}(f_Y - \bar{f}_Y)$ (Cheng, et al. (2012))

Modelo de variable latente

Para las variables ordinales $\mathbf{X}$...

- p variables latentes $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_p)^T$ continuas
- conjunto de umbrales $\boldsymbol{\theta}^{(j)}$ para cada $j = 1, 2, \ldots, p$

$$-\infty = \theta_0^{(j)} < \theta_1^{(j)} < \cdots < \theta_{K_j-1}^{(j)} < \theta_{G_j}^{(j)} = +\infty$$

Modelo de variable latente

Para las variables ordinales $\mathbf{X}$...

- p variables latentes $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_p)^T$ continuas
- conjunto de umbrales $\boldsymbol{\theta}^{(j)}$ para cada $j = 1, 2, \ldots, p$

$$-\infty = \theta_0^{(j)} < \theta_1^{(j)} < \cdots < \theta_{K_j-1}^{(j)} < \theta_{G_j}^{(j)} = +\infty$$

- $\mathbf{X}_j = g \iff Z_j \in [\theta_{g-1}^{(j)}, \theta_g^{(j)})$

Modelo de variable latente

Para las variables ordinales $\mathbf{X}$...

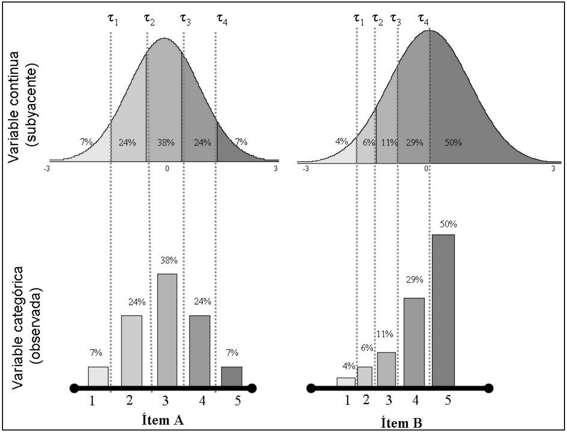
- p variables latentes $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_p)^T$ continuas
- conjunto de umbrales $\boldsymbol{\theta}^{(j)}$ para cada $j = 1, 2, \ldots, p$

$$-\infty = \theta_0^{(j)} < \theta_1^{(j)} < \cdots < \theta_{K_j-1}^{(j)} < \theta_{G_j}^{(j)} = +\infty$$

- $\mathbf{X}_j = g \iff Z_j \in [\theta_{g-1}^{(j)}, \theta_g^{(j)})$
- $\boldsymbol{\Theta} \doteq \{\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(p)}\} = \{\theta_0^{(1)}, \ldots, \theta_{G_1}^{(1)}, \ldots\ldots, \theta_0^{(p)}, \ldots, \theta_{G_p}^{(p)}\}$

Modelo de variable latente

$$\Pr(X_j = g|Y) = \Pr(\theta_{g-1}^{(j)} \leq Z_j < \theta_g^{(j)}|Y)$$



Modelo normal multivariado

- $\mathbf{V} \doteq \begin{bmatrix} \mathbf{Z} \\ \mathbf{W} \end{bmatrix}$

Modelo normal multivariado

- $\mathbf{V} \doteq \begin{bmatrix} \mathbf{Z} \\ \mathbf{W} \end{bmatrix}$
- $\mathbf{V}|\mathbf{H}, Y \sim N(\boldsymbol{\mu} + \boldsymbol{\Delta}\boldsymbol{\alpha}\boldsymbol{\xi}(f_Y - \bar{f}_Y) + \boldsymbol{\beta}(\mathbf{H} - \boldsymbol{\mu}_{\mathbf{H}}), \boldsymbol{\Delta})$:

$$
\begin{aligned}
f(\mathbf{V}|\mathbf{H}, Y) = (2\pi)^{-\frac{t}{2}}|\boldsymbol{\Delta}|^{-\frac{1}{2}} \\
\exp\Big\{ -\frac{1}{2}\Big((\mathbf{V} - \boldsymbol{\mu}) - \boldsymbol{\Delta}\boldsymbol{\alpha}\boldsymbol{\xi}(f_Y - \bar{f}_Y) - \boldsymbol{\beta}(\mathbf{H} - \boldsymbol{\mu}_{\mathbf{H}})\Big)^T \\
\boldsymbol{\Delta}^{-1}\Big((\mathbf{V} - \boldsymbol{\mu}) - \boldsymbol{\Delta}\boldsymbol{\alpha}\boldsymbol{\xi}(f_Y - \bar{f}_Y) - \boldsymbol{\beta}(\mathbf{H} - \boldsymbol{\mu}_{\mathbf{H}})\Big)\Big\}
\end{aligned}
$$

Reducción suficiente con predictores mixtos

$$
\begin{aligned}
f(\mathbf{V},\mathbf{H}|Y) = (2\pi)^{-\frac{t}{2}}|\boldsymbol{\Delta}|^{-\frac{1}{2}} \\
\exp\Big\{ -\frac{1}{2}\Big((\mathbf{V}-\boldsymbol{\mu}) - \boldsymbol{\Delta}\boldsymbol{\alpha}\boldsymbol{\xi}(f_Y - \bar{f}_Y) - \boldsymbol{\beta}(\mathbf{H}-\boldsymbol{\mu}_{\mathbf{H}})\Big)^T \\
\boldsymbol{\Delta}^{-1}\Big((\mathbf{V}-\boldsymbol{\mu}) - \boldsymbol{\Delta}\boldsymbol{\alpha}\boldsymbol{\xi}(f_Y - \bar{f}_Y) - \boldsymbol{\beta}(\mathbf{H}-\boldsymbol{\mu}_{\mathbf{H}})\Big) \\
+ \operatorname{vech}^T(\mathbf{H}\mathbf{H}^T)[\boldsymbol{\tau}_0 + \boldsymbol{\tau}\boldsymbol{\kappa}(f_Y - \bar{f}_Y)] - \log(G(\boldsymbol{\Gamma}^Y))\Big\}
\end{aligned}
$$

Reducción suficiente con predictores mixtos

$$f(\mathbf{V},\mathbf{H}|Y) = (2\pi)^{-\frac{t}{2}}|\boldsymbol{\Delta}|^{-\frac{1}{2}}$$
$$\exp\Big\{-\frac{1}{2}\big((\mathbf{V}-\boldsymbol{\mu})-\boldsymbol{\Delta\alpha\xi}(f_Y-\bar{f}_Y)-\boldsymbol{\beta}(\mathbf{H}-\boldsymbol{\mu}_{\mathbf{H}})\big)^T$$
$$\boldsymbol{\Delta}^{-1}\big((\mathbf{V}-\boldsymbol{\mu})-\boldsymbol{\Delta\alpha\xi}(f_Y-\bar{f}_Y)-\boldsymbol{\beta}(\mathbf{H}-\boldsymbol{\mu}_{\mathbf{H}})\big)$$
$$+\ \mathrm{vech}^T(\mathbf{H}\mathbf{H}^T)[\boldsymbol{\tau}_0+\boldsymbol{\tau\kappa}(f_Y-\bar{f}_Y)]-\log(G(\boldsymbol{\Gamma}^Y))\Big\}$$

Familia exponencial con:

$$\mathbf{T}(\mathbf{V},\mathbf{H}) = \begin{pmatrix} \mathbf{V} \\ \mathbf{H} \\ \mathrm{vech}(\mathbf{H}\mathbf{H}^T) \end{pmatrix} \qquad \boldsymbol{\eta}_Y-\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\alpha\xi} \\ -\boldsymbol{\beta}^T\boldsymbol{\alpha\xi} \\ \boldsymbol{\tau\kappa} \end{pmatrix}(f_Y-\bar{f}_Y),$$

Reducción suficiente con predictores mixtos

Bura, Duarte and Forzani (2015)

$$\mathbf{R}(\mathbf{V},\mathbf{H}) = \mathbf{a}^T(\mathbf{T}(\mathbf{V},\mathbf{H}) - \mathrm{E}(\mathbf{T}(\mathbf{V},\mathbf{H}))),$$
$$\mathbf{a} = \mathrm{span}\{\boldsymbol{\eta}_Y - \boldsymbol{\eta}, Y \in S_Y\}$$

Reducción suficiente con predictores mixtos

Bura, Duarte and Forzani (2015)

$$\mathbf{R}(\mathbf{V},\mathbf{H}) = \mathbf{a}^T(\mathbf{T}(\mathbf{V},\mathbf{H}) - \mathrm{E}(\mathbf{T}(\mathbf{V},\mathbf{H}))),$$
$$\mathbf{a} = \mathrm{span}\{\boldsymbol{\eta}_Y - \boldsymbol{\eta}, Y \in S_Y\}$$

Teorema

Una reducción suficiente para la regresión $Y|(\mathbf{X},\mathbf{W},\mathbf{H})$ *es*

$$\mathbf{R}(\mathbf{X},\mathbf{W},\mathbf{H}) = \mathbf{a}^T(\mathbf{T}(\mathbf{X},\mathbf{W},\mathbf{H}) - \mathrm{E}(\mathbf{T}(\mathbf{X},\mathbf{W},\mathbf{H})))$$

donde $\mathbf{a} = \mathrm{span}\{\boldsymbol{\eta}_Y - \bar{\boldsymbol{\eta}}, Y \in S_Y\} = (\boldsymbol{\alpha}^T, -\boldsymbol{\alpha}^T\boldsymbol{\beta}, (\boldsymbol{\xi}\boldsymbol{\xi}^T)^{-1}\boldsymbol{\xi}\boldsymbol{\iota}^T\boldsymbol{\kappa}^T)^T$.

Estimación: PFCmix

$$\{(\mathbf{X}_i, \mathbf{W}_i, \mathbf{H}_i, Y_i)\}_{i=1}^{n}$$

$$\Downarrow$$

$$\hat{\boldsymbol{\tau}}_0, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Delta}}, \hat{\boldsymbol{\beta}}$$

Estimación: PFCmix

$$\{(\mathbf{X}_i, \mathbf{W}_i, \mathbf{H}_i, Y_i)\}_{i=1}^{n}$$

$$\Downarrow$$

$$\hat{\boldsymbol{\tau}}_0, \hat{\boldsymbol{\tau}}$$

- Ising con covariables (Cheng, et al. 2012)

Estimación: PFCmix

$$\{(\mathbf{X}_i, \mathbf{W}_i, \mathbf{H}_i, Y_i)\}_{i=1}^{n}$$

$$\Downarrow$$

$$\hat{\boldsymbol{\tau}}_0, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\Theta}}$$

- Ising con covariables (Cheng, et al. 2012)
- Método iterativo:
  - Para cada $\boldsymbol{\Theta}^{(j)}$, resolvemos $L_g(\boldsymbol{\theta}) = 0$:

$$L_{\boldsymbol{g}}(\boldsymbol{\theta}) \doteq \#\{i : [\mathbf{X}_j]_i \leq \boldsymbol{g}\} - \sum_{i=1}^{n} \Phi\left(\frac{\boldsymbol{\theta} - [\boldsymbol{\Delta}\boldsymbol{\alpha}\boldsymbol{\xi}]_j (f_{Y_i} - \bar{f}_Y) - \boldsymbol{\beta}_j (\mathbf{H}_i - \boldsymbol{\mu}_{\mathbf{H}})}{[\boldsymbol{\Delta}]_{jj}}\right)$$

Estimación: PFCmix

$$\{(\mathbf{X}_i, \mathbf{W}_i, \mathbf{H}_i, Y_i)\}_{i=1}^n$$

$$\Downarrow$$

$$\hat{\boldsymbol{\tau}}_0, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Delta}}, \hat{\boldsymbol{\beta}}$$

- Ising con covariables (Cheng, et al. 2012)
- Método iterativo:
  - Para cada $\boldsymbol{\Theta}^{(j)}$, resolvemos $L_g(\boldsymbol{\theta}) = 0$:

$$L_{\boldsymbol{g}}(\boldsymbol{\theta}) \doteq \#\{i : [\mathbf{X}_j]_i \leq \boldsymbol{g}\} - \sum_{i=1}^{n} \Phi\left(\frac{\boldsymbol{\theta} - [\boldsymbol{\Delta\alpha\xi}]_j (f_{Y_i} - \bar{f}_Y) - \boldsymbol{\beta}_j (\mathbf{H}_i - \boldsymbol{\mu}_{\mathbf{H}})}{[\boldsymbol{\Delta}]_{jj}}\right)$$

  - Algoritmo **EM**

Aplicación usando EPH-Argentina

Aplicación usando EPH-Argentina

- EPH 3er. trimestre de 2013 (INDEC).

Aplicación usando EPH-Argentina

- EPH 3er. trimestre de 2013 (INDEC).
- Predictores: 8 ordinales, 4 binarios y 2 continuas.

Aplicación usando EPH-Argentina

- EPH 3er. trimestre de 2013 (INDEC).
- Predictores: 8 ordinales, 4 binarios y 2 continuas.
- Respuesta: dicotómica (pobreza)

Aplicación usando EPH-Argentina

- EPH 3er. trimestre de 2013 (INDEC).
- Predictores: 8 ordinales, 4 binarios y 2 continuas.
- Respuesta: dicotómica (pobreza)
- Heterogeneidad regional $\rightsquigarrow$ Estimación de diferentes I para 5 regiones (GBA, Pampeana, NOA, NEA y Patagonia).

Aplicación usando EPH-Argentina

- EPH 3er. trimestre de 2013 (INDEC).
- Predictores: 8 ordinales, 4 binarios y 2 continuas.
- Respuesta: dicotómica (pobreza)
- Heterogeneidad regional $\rightsquigarrow$ Estimación de diferentes I para 5 regiones (GBA, Pampeana, NOA, NEA y Patagonia).
- Comparamos con:
  - PFCord
  - NLPCA

- Predictores ordinales:
    - Ubicación del agua potable (3 cat.);
    - Forma de obtención del agua potable (3 cat.);
    - Tipo de baño (3 cat.);
    - Desage del baño (4 cat.),
    - Forma de compartir el baño (3 cat.)
    - Calidad de la Vivienda (4 cat);
    - Combustible predominante para cocinar alimentos (3 cat.);
    - Escolaridad del Jefe/a de hogar (7 cat.).
- Predictores binarios:
    - Actividad del jefe/a (Ocupado o no);
    - Vivienda cercana a basurales (si-no);
    - Vivienda ubicada en zona inundable (si-no);
    - Vivienda ubicada en villa de emergencia (si-no).
- Predictores continuos:
    - Hacinamiento: ratio entre ambientes de la vivienda y cantidad de miembros del hogar;
    - Horas trabajadas por el jefe/a en la ltima semana.

MSE promediado en 10-fold validación

Método	Variables	Bs. As.	Pampa	Regiones NOA	NEA	Patagonia
PFCMix	X,H,W	**0.2131**	**0.1881**	**0.2809**	**0.3064**	**0.1321**
PFCOrd	X	0.2164	0.1921	0.3292	0.3947	0.1322
PFCOrd	X,H	0.2185	0.1909	0.3309	0.3851	0.1321
NLPCA	X,H,W	0.2335	0.1940	0.3331	0.3686	0.1324

MUCHAS GRACIAS!

FIQ UNL

UNIVERSIDAD NACIONAL DEL LITORAL
FACULTAD DE INGENIERÍA QUÍMICA

FIQUNL
www.fiq.unl.edu.ar