

Midiendo la Vulnerabilidad a la Pobreza Multidimensional con Clasificadores de Redes Bayesianas

Mauricio Gallardo

Universidad Católica del Norte



¿QUÉ ES LA VULNERABILIDAD A LA POBREZA?

- ▶ Un concepto controversial.
- ▶ Una distinción fundamental:
 - ▶ *pobreza*: realización efectiva de un nivel de bienestar por debajo de cierto umbral mínimo ('línea de pobreza')
 - ▶ *vulnerabilidad*: incertidumbre de experimentar pobreza (riesgo de ser pobre).
- ▶ Una definición intuitiva: 'estar en riesgo de caer en pobreza cuando no se es pobre, o en riesgo de continuar siendo pobre cuando ya se es pobre'.
- ▶ Vulnerabilidad implica una 'perspectiva ex ante': antes que se revele la incertidumbre. Sin importar si eres pobre o no-pobre 'ex-post'.
- ▶ Se argumenta que la vulnerabilidad es 'forward-looking', mientras que la pobreza es 'backwards looking'.

VULNERABILIDAD A LA POBREZA DE INGRESOS:

PRINCIPALES MODELOS

- ▶ Una literatura muy extensa (Gallardo, 2018 - JOES).
- ▶ Algunos modelos:
 - ▶ Línea de pobreza extendida (Cafiero and Vakis, 2006):

$$y_j < z^v, z^v > z$$

- ▶ Pobreza esperada (Chaudhuri et al., 2002):

$$V_{j,t} = F(z) = \Pr(y_{j,t+1} \leq z)$$

- ▶ Baja utilidad esperada (Ligon and Schechter, 2003):

$$V_j = U_j(z) - E[U_j(y_j)]$$

- ▶ Utilidad esperada censurada (Calvo and Dercon, 2013):

$$V_j = 1 - E\left(\frac{\text{Min}\{y_j, z\}}{z}\right)^\rho, 0 < \rho < 1$$

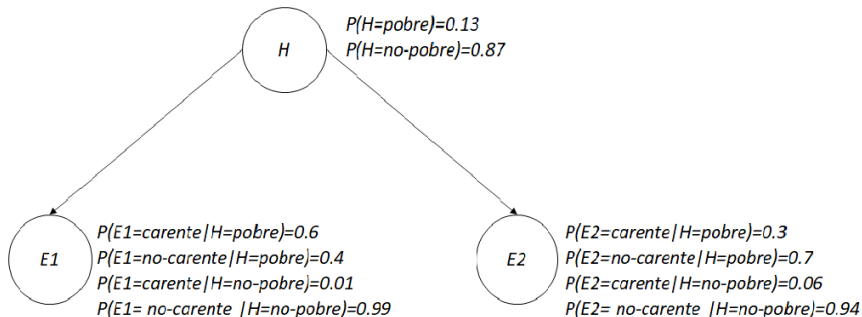
- ▶ Enfoque seimétrico de 'media-riesgo' (Gallardo, 2013).

PROBLEMA DE MEDICIÓN MULTIDIMENCIONAL

- ▶ Problema complejo: VPM implica modelar la incertidumbre y el riesgo multidimensional.
- ▶ Soluciones vigentes son de dos tipos:
 - i) Replicación de una medida de pobreza (sin modelo de incertidumbre), ej., OPHI usa un enfoque de umbral de pobreza extendido con $k \in [20 - 33\%]$.
 - ii) Método que enfrenta el problema de estimación multidimensional transformándolo es uno o varios problemas unidimensionales, ej., Feeny y McDonald (2016), *Probit Heterocedástico*; Gallardo (2020), *Regresión Logística Multinivel* o *Probit Multinivel*.
- ▶ Tanto Feeny y McDonald (2016) como Gallardo (2020) usan métodos unidimensionales para enfrentar un desafío multidimensional.

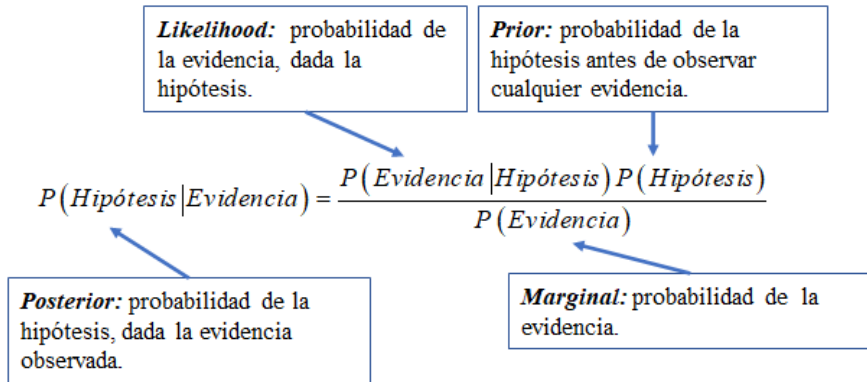
¿QUÉ ES UNA RED BAYESIANA?

- $B(G, \Theta)$: modelo probabilístico multivariado que se define a través de dos elementos, un gráfico acíclico dirigido G y un vector de parámetros de probabilidad Θ .
- Un Ejemplo:



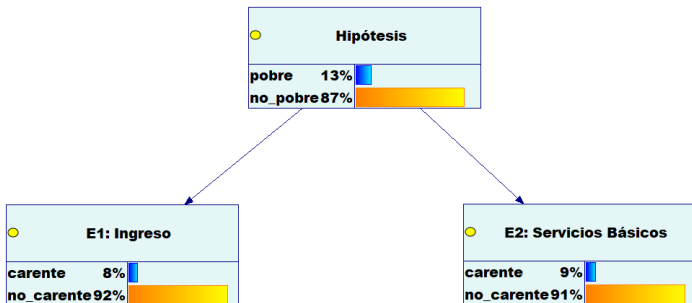
TEOREMA DE BAYES

La Red Bayesiana permite inferir la probabilidad de la hipótesis dada la evidencia a través del *Teorema de Bayes*.



'PRIORS' DEL EJEMPLO

- En nuestro ejemplo asumimos la regla de que un hogar es multidimensionalmente pobre si tiene carencias en ambas dimensiones. La frecuencia relativa de pobreza en la población es 0,13. Esta será nuestro **Prior** para la hipótesis de ser multidimensionalmente pobre, mientras que las **Likelihood** son las probabilidades condicionales presentadas antes: $P(E1|H)$ y $P(E2|H)$.

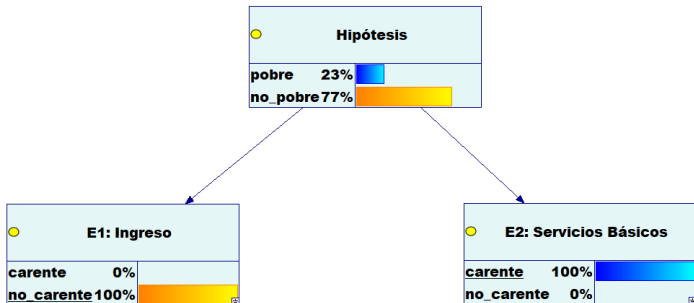


IMPLICANCIAS DE POLÍTICA PÚBLICA: HOGAR A

- Evidencia: carente en servicios básicos, no carente en ingresos. ¿Cómo esta evidencia cambia nuestra creencia sobre H ?
- Usando el Teorema de Bayes:

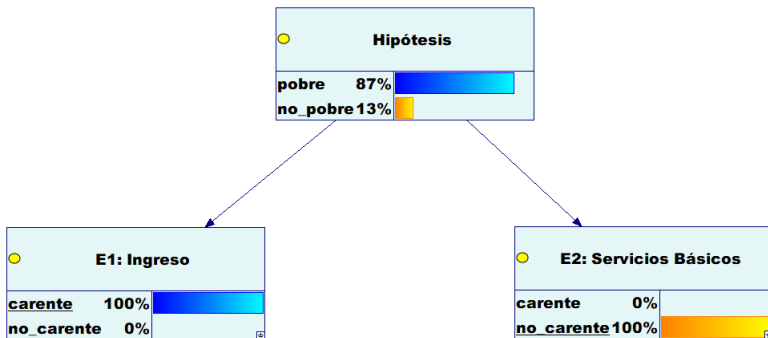
$$P(H = 0|E_1 = 1, E_2 = 0) = \frac{P(E_1 = 1, E_2 = 0|H = 0)P(H = 0)}{P(E_1 = 1, E_2 = 0)}$$

$$P(H = 0|E_1 = 1, E_2 = 0) = \frac{(0.4 \times 0.3)0.13}{(0.4 \times 0.3)0.13 + (0.99 \times 0.06)0.87} = 0.23$$



IMPLICANCIAS DE POLÍTICA PÚBLICA: HOGAR B

- Evidencia: carente en ingresos, no carente en servicios básicos. ¿Cómo esta evidencia cambia nuestra creencia sobre H ?
- Usando el Teorema de Bayes, vemos que la probabilidad de la hipótesis dada esta evidencia eleva nuestra creencia de esta probabilidad a 0.87:



OBJETIVO DE INVESTIGACIÓN Y DATOS

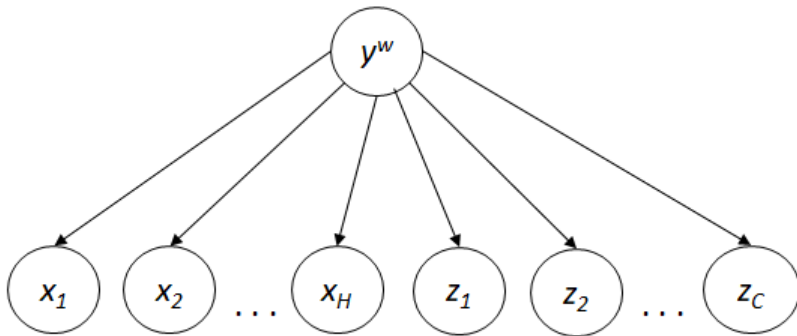
- ▶ Estimar las medidas VPM de Feeny y McDonald (2016) y de Gallardo (2020) usando métodos multidimensionales MBC.
- ▶ Extender el conocimiento de las medidas VPM hacia un nuevo enfoque que integre las ventajas de las metodologías de Feeny y McDonald (2016) y Gallardo (2020).
 - ▶ Enfoque de Feeny y McDonald (2016) se funda en una estimación de la probabilidad agregada.
 - ▶ Enfoque de Gallardo (2020) se funda en una estimación de las probabilidades de cada uno de los componentes.
 - ▶ MBC permite estimar todas las probabilidades simultáneamente.
- ▶ Datos: CASEN-2017 para las características del hogar, CENSO-2017 para las características de la comuna.

¿QUÉ QUEREMOS ESTIMAR?

- ▶ Estructura informacional:
 - ▶ y^w : variable binaria igual a uno si la persona es no-pobre multidimensional e igual a cero en otro caso.
 - ▶ $y_i = (y_{i1}, \dots, y_{iM})$: vector de logros compuesto de variables binarias iguales a uno cuando la persona es no-carente en la dimensión (indicador de bienestar) i , e igual a cero en otro caso.
 - ▶ $x_i = (x_{i1}, \dots, x_{iH})$: vector de características del hogar. Todas variables aleatorias categóricas.
 - ▶ $z_i = (z_{i1}, \dots, z_{iC})$: vector de características de la comuna. Todas variables aleatorias categóricas.
- ▶ Nuestro objetivo es estimar las siguientes funciones usando Clasificadores de Redes Bayesianas:
 1. $h_1 : (x_{i1}, \dots, x_{iH}, z_{i1}, \dots, z_{iC}) \rightarrow P(y_i^w | x_i, z_i)$.
 2. $h_2 : (x_{i1}, \dots, x_{iH}, z_{i1}, \dots, z_{iC}) \rightarrow P(y_{im} | x_i, z_i)$
 3. $(y_{i1}, \dots, y_{iM}, x_{i1}, \dots, x_{iH}, z_{i1}, \dots, z_{iC}) \rightarrow (P(y_i^w | y_i), P(y_{im} | x_i, z_i))$

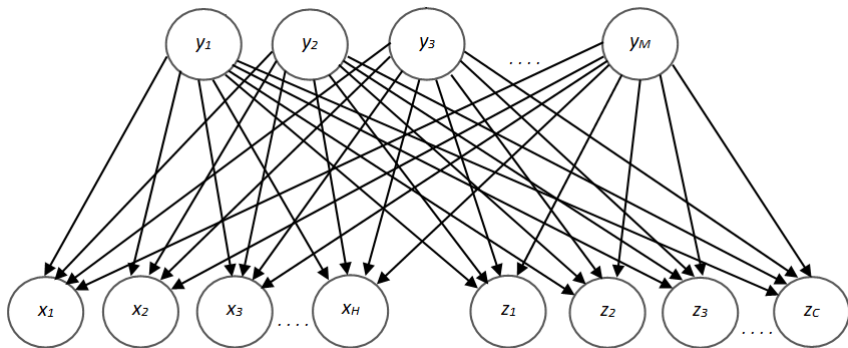
MODELO 1: CLASIFICADOR BAYESIANO INGENUO

De acuerdo a su tipología la variable clase y^w es el 'padre' de las variables de características representadas por los elementos de los vectores aleatorios x y z .



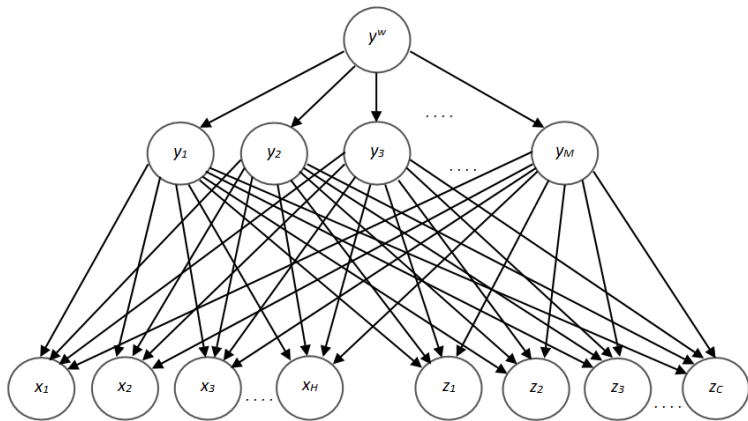
MODELO 2: MBC DE DOS NIVELES

Un MBC es una RB de topología específica en donde las variables clase son los 'padres' (variables condicionantes) de las variables de características.



MODELO 3: MBC DE TRES NIVELES

Este MBC difiere del anterior solo porque incorpora una variable 'super-clase' que aparece como el primer ancestro en su topología.



REGLAS DE IDENTIFICACIÓN

- ▶ Enfoque FM:
 - ▶ Vulnerable: $P(y_i^w = 0) > H$ (in Chile 2017 $H = 0.207$).
 - ▶ Altamente vulnerable: $P(y_i^w = 0) > 0.5$.
 - ▶ Relativamente vulnerable: $H < P(y_i^w = 0) < 0.5$.
- ▶ Enfoque Gallardo (2020):
 - ▶ Vulnerable dimensional: $\mu_{im} - \gamma_m \sigma_{im}^- < z^m$.
 - ▶ Vulnerable multidimensional: $s_i^y \geq k$, (in Chile $k = 0.225$).
- ▶ Nuevo enfoque MBC:
 - ▶ Vulnerable dimensional: $\mu_{im} - \gamma_m \sigma_{im}^- \leq 0.5$.
 - ▶ Vulnerable multidimensional: $\mu_i - \gamma \sigma_i^- \leq 0.5$.
 - ▶ Vulnerable inducido por pobreza: $\mu_i \leq 0.5$.
 - ▶ Vulnerable inducido por riesgo: $\mu_i > 0.5$ and $\mu_i - \gamma \sigma_i^- \leq 0.5$
- ▶ Enfoque FM-A: similar al nuevo MBC pero estimado con el Modelo 1 (CBI).

IPM DE REFERENCIA

Educación 22.5%	Salud 22.5%	Empleo y seguridad social 22.5%	Vivienda y ambiente 22.5%	Redes y cohesion social 10%
Asistencia escolar de niños 7.5%	Malnutrición infantil 7.5%	Empleo 7.5%	Habitabilidad de la vivienda 7.5%	Participación social y redes de apoyo 3.33%
Rezago escolar 7.5%	Afiliación a Seguro de salud 7.5%	Seguridad social 7.5%	Servicios Básicos 7.5%	Tratamiento igualitario 3.33%
Escolaridadde adultos 7.5%	Acceso a atención de salud 7.5%	Pensión de retiro 7.5%	Ambiente de la vivienda 7.5%	Seguridad 3.33%

EVALUACIÓN DE DESEMPEÑO DEL MBC (5-FOLD)

Dimension	Model 1	Model 2	Model 3
Overall accuracy	0.80	0.84	0.96
Average Accuracy over Dimensions	--	0.85	0.86
<i>Accuracy by dimensions</i>			
Children's school attendance		0.92	0.92
Schooling lag		0.92	0.92
Years of adults' schooling		0.68	0.73
Child malnutrition		0.89	0.90
Health insurance affiliation		0.90	0.90
Access to healthcare		0.91	0.91
Employment		0.87	0.87
Social security		0.66	0.70
Retirement insurance		0.84	0.85
Habitability		0.78	0.80
Basic services		0.85	0.88
Local environment		0.85	0.86
Equal treatment		0.85	0.85
Safety		0.87	0.87
Social participation and support		0.90	0.90

ANÁLISIS DE SENSIBILIDAD

<i>Different multidimensional poverty cut-offs</i>				
Bayesian classifier	<i>k=0.225</i>	<i>k=0.15</i>	<i>k=0.3</i>	<i>Range</i>
Model 1	0.80	0.68	0.92	0.231
Model 2	0.84	0.75	0.84	0.098
Model 3	0.96	0.99	0.98	0.024
<i>Different weight configurations</i>				
	<i>Official MPI</i>	<i>Equal weights</i>	<i>Lower weights for networks and social cohesion blocks</i>	<i>Range</i>
Model 1	0.80	0.88	0.80	0.084
Model 2	0.84	0.91	0.82	0.092
Model 3	0.96	0.97	0.96	0.009
<i>Different MPI compositions</i>				
	<i>Official MPI</i>	<i>Four dimensional blocks MPI</i>	<i>Official MPI plus monetary dimension</i>	<i>Range</i>
Model 1	0.80	0.80	0.84	0.044
Model 2	0.84	0.81	0.86	0.048
Model 3	0.96	0.97	0.98	0.013

PRINCIPALES MEDIDAS VPM

VMP measurement	<i>Headcount ratio</i>	<i>Poverty- induced vulnerability</i>	<i>Risk- induced vulnerability</i>	<i>Vulnerability to poverty ratio</i>
FM	0.31 (0.0027)	0.14 (0.0019)	0.17 (0.0023)	1.50
FM-A (gamma=1)	0.36 (0.0025)	0.14 (0.0019)	0.23 (0.0024)	1.75
GM (gamma=0.8)	0.37 (0.003)	0.09 (0.002)	0.28 (0.0048)	1.79
MBC (gamma=1)	0.33 (0.0029)	0.17 (0.0023)	0.16 (0.0023)	1.59

CONCLUSIONES

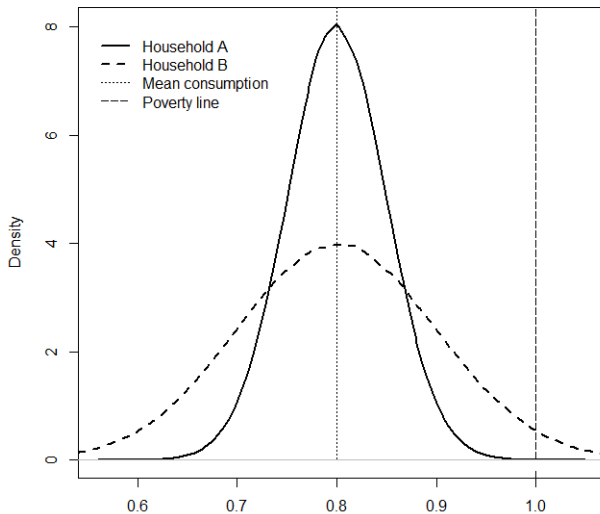
- ▶ Las estrategias de estimación de MBC ofrecen una buena alternativa a los modelos tradicionales Probit y Logit para predecir la probabilidad de ser multidimensionalmente pobre.
- ▶ Los MBC son estrategias más adecuadas porque son un procedimiento multidimensional para resolver un problema de estimación que es de naturaleza multidimensional.
- ▶ El rendimiento predictivo de los modelos de clasificadores bayesianos es alto y, por lo tanto, son adecuados para ser aplicados por los responsables de la formulación de políticas públicas en la práctica.
- ▶ Aunque hemos demostrado que el MBC se puede utilizar para estimar las mediciones de FM y GM VPM, también introdujimos un nuevo enfoque que puede integrar las ventajas de los enfoques anteriores en un marco analítico unificado.

Midiendo la Vulnerabilidad a la Pobreza Multidimensional con Clasificadores de Redes Bayesianas

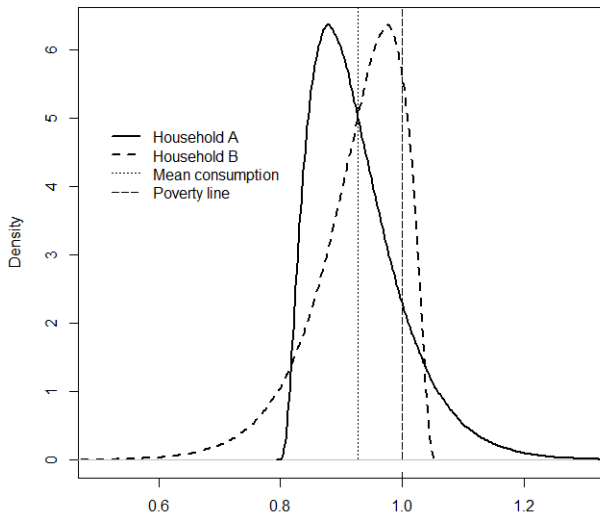
Mauricio Gallardo
Universidad Católica del Norte



RELEVANCIA DE LA SENSIBILIDAD AL RIESGO



RELEVANCIA DE LA SENSIBILIDAD A LA ASIMETRÍA



MODELO DE COMPORTAMIENTO DE MEDIA-RIESGO

Regla de dominancia en media-riesgo:

$$y_i \succ y_j \Leftrightarrow \mu_i \geq \mu_j \wedge r_i \leq r_j$$

$$y_i \succ y_j \Leftrightarrow \mu_i - \gamma r_i > \mu_j - \gamma r_j$$

Un parámetro de riesgo apropiado (semi-desviación a la baja):

$$\sigma_i^- = \sqrt{E \left\{ \min [(y_i - \mu_i), 0]^2 \right\}}$$

Semi-desviación a la baja para variables binarias:

$$\sqrt{\widehat{p}_i^2 (1 - \widehat{p}_i)}$$