Esquema del IPM-SAE-CEPAL
24 de febrero de 2022

Supondremos, para simplificar el ejemplo, que el indicador de interés está compuesto por tres variables: una disponible en el censo (*y_1*) y dos que deben ser predichas para cada unidad en el censo. Además, de las dos variables no observadas, una es discreta dicotómica (*y_2*) y la otra es una función del ingreso laboral (*y_3*).

Por la naturaleza de estas últimas dos variables, el enfoque de modelación, estimación y predicción deberá adaptarse apropiadamente al estado de la cuestión de los modelos. Esto es, la variable *y_2* será modelada a través de un modelo EPP (*empirical plug-in*), mientras que la variable *y_3* será modelada con un EBP (*empirical best/bayes*).

En primera instancia, cada uno de los dos modelos se ajusta en la encuesta, usando los datos disponibles en BADEHOG. A partir de estos modelos tendremos estimaciones para los parámetros de regresión y para los efectos aleatorios a nivel de municipio (obviamente, diferentes para cada modelo 🤣). Al contrario de lo que se plantea en los ejercicios regulares de SAE para ingresos, pobreza o cualquier otro indicador individual, nuestro propósito no es obtener medidas de resumen predictivas sobre cada variable *y_2, y_3*, sino utilizar su predicción, en conjunto con la variable observada *y_1* para, a su vez, obtener predicciones de la privación del individuo en la dimensión.

El modelo EPP afirma que *y_2* sigue la siguiente distribución de probabilidad

$$y_{di} \mid y_{ds} \sim N(\mu_{di|s}, \sigma^2_{di|s})$$

El modelo EBP afirma que *y_3* sigue la siguiente distribución de probabilidad

$$y_{di} \mid u_d \sim Ber(p_{di})$$

Por lo tanto, utilizando un enfoque paramétrico, haremos uso de los métodos de simulación para obtener estas predicciones, no solo una vez, sino muchas veces. De esta forma, cada iteración del Monte Carlo EPP inducirá una predicción para *y_2* y cada iteración del Monte Carlo EBP inducirá una predicción para *y_3*. Así se vería la base de datos super-ampliada para las J iteraciones de Monte Carlo:

K	y_{-1}	$y_{-2}^{(1)}$	$y_2^{(2)}$	…	$y_2^{(J)}$	$y_3^{(1)}$	$y_3^{(2)}$	…	$y_3^{(J)}$
1	1	1	0	…	1	0	1	…	1
2	1	0	1	…	1	1	1	…	1
⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋱	⋮
N	0	1	0	…	0	0	1		0

Los colores nos sirven para enfatizar que cada bloque (verde y azul) proviene de una medida de probabilidad diferente (EPP y EBP, respectivamente). Ahora, la variable y_1, junto con la primera iteración de los Monte Carlos dará origen a una predicción del IPM para todos los individuos. Así sucesivamente para las J realizaciones de la anterior gráfica, tendríamos la siguiente estructura para el IPM:

K	$IPM^{(1)}$	$IPM^{(2)}$	…	$IPM^{(J)}$	$E(IPM)$
1	1	0	…	1	0.87
2	1	1	…	1	0.23
⋮	⋮	⋮	⋱	⋮	⋮
N	1	0	…	0	0.04

¡¡¡¡ACÁ ESTÁ LA GRAN CONTRIBUCIÓN DEL PAPER!!!!

Utilizando este conjunto de datos simulados, fácilmente se llegan a las estimaciones municipales del IPM, utilizando la esperanza como predictor (de forma similar a un EPP). Acá valdría la pena verificar si es que este IPM puede verse como una realización de una variable aleatoria Bernoulli, puesto que al final cada individuo en el censo tendrá una probabilidad de ser pobre multidimensional.

La estimación de ECM debería seguir los principios de Gonzáles-Manteiga et. al. (2007), que podría ser resumido de la siguiente manera:

1. Ajustar el EPP y el EBP a los datos de la muestra, obteniendo todos los estimadores necesarios.
2. Generar efectos aleatorios EPP, efectos aleatorios EBP y errores EBP.
3. Generar dos pseudo-censos Bootstrap, uno para *y_1* y otro para *y_2*.
4. Seleccionar una muestra Bootstrap, siguiendo los lineamientos de la muestra original.
5. Ajustar modelos EBP y EPP, apropiadamente a los pseudo-censos y calcular los predictores del IPM.
6. Repetir B veces y calcular la estimación del ECM.
7. Cruzar los dedos para que esto se ejecute rápido.

¡¡¡Y listo!!!