

Elección discreta: Teoría, estimación y métodos numéricos

Tarea 4: Mixed logit

El Colegio de México

Prof. Edwin Muñoz Rodríguez

Alumno: Max Brando Serna Leyva

Índice

1. Instrucciones iniciales	2
2. Inciso 1	3
3. Inciso 2	4
4. Inciso 3	5
5. Inciso 4	6

1. Instrucciones iniciales

En esta tarea implementaremos el Mixed logit en nuestra base de datos `yogurt.csv`.

Considere el siguiente modelo de utilidad aleatoria:

$$u_{jm} = \alpha_{jm} + \beta_{mp}price_{jm} + \beta_{mf}feat_{jm} + \varepsilon_{jm} \quad (1.1)$$

donde $\varepsilon_{jm} \sim Gumbel(0, 1)$ son i.i.d. entre consumidores y productos. Normalice $\alpha_4 = 0$.

Asuma que los coeficientes α_{jm} , $j = 1, 2, 3$ así como β_{mp}, β_{mf} son aleatorios, distribuidos normales con medias $\bar{\alpha}_j, \bar{\beta}_p, \bar{\beta}_f$ y matriz de varianzas y covarianzas

$$\Sigma_{5 \times 5} = \begin{bmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & & \\ & & \sigma_3^2 & & \\ & & & \sigma_p^2 & \\ & & & & \sigma_f^2 \end{bmatrix}$$

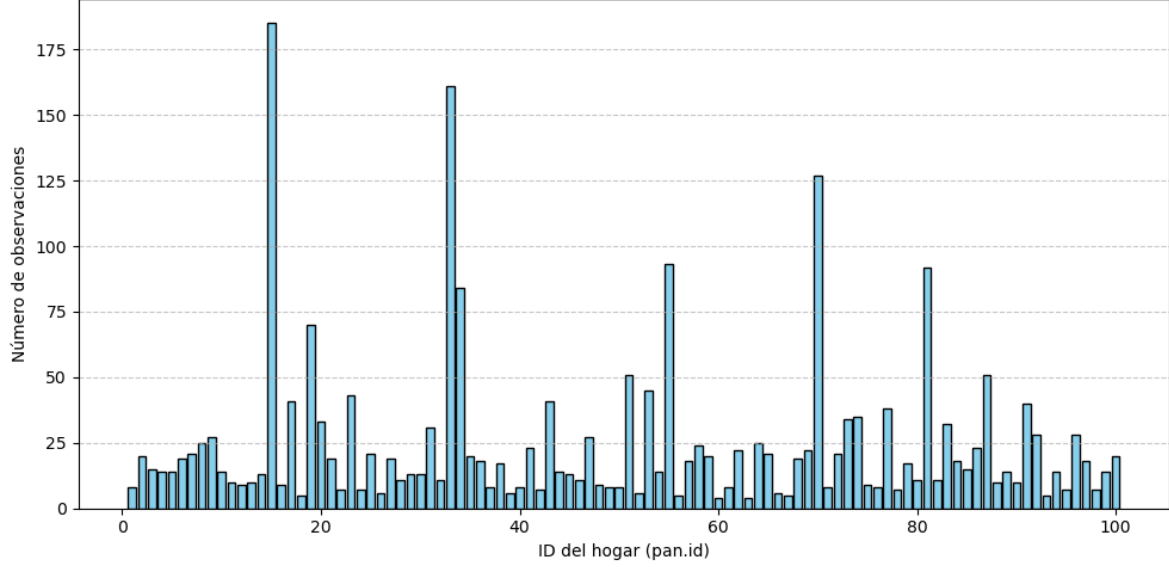
con todos los términos fuera de la diagonal iguales a cero. Esto es, son independientes.

1. Haga un supuesto sobre la estabilidad de las preferencias de los hogares, y asegúrese de que su modelo sea consistente con dicho supuesto. Para esto, recuerde que la variable `pan.id` es un identificador de hogar.
2. Escriba claramente el modelo que estima, la verosimilitud y la verosimilitud simulada.
3. Para asegurarse de que $\Sigma_{5 \times 5}$ es definida positiva implemente la descomposición de Cholesky.
4. Reporte las estimaciones y errores estándar para todos los parámetros con $R = 50$, usando método de montecarlo. Discuta sus resultados.

2. Inciso 1

Revisando la variable `pan.id` nos damos cuenta de que un mismo hogar entra a la base de datos en repetidas ocasiones, lo que significa que **tenemos más de una situación de elección para cada hogar**. La figura 1 muestra un histograma donde se observa el número de situaciones de elección para cada hogar en la muestra.

Figura 1: Histograma de `pan.id`



Esto plantea un problema en cuanto a la caracterización de las preferencias.

Sea $b_n = (\alpha_{n1}, \alpha_{n2}, \alpha_{n3}, \beta_{np}, \beta_{nf})^\top$, con $n = 1, 2, \dots, N$, donde N es el número de hogares en la muestra. Si suponemos que en cada situación de elección, tenemos coeficientes distintos, entonces tendríamos que estimar b_{n,t_n} , donde t_n sería un subíndice para identificar la situación de elección t del individuo n . **Para evitar esto, haremos un supuesto sobre la estabilidad de las preferencias**, en el que para cada individuo n , b_n es constante en toda situación de elección t_n ; es decir, vamos a suponer que las preferencias de cada individuo son estables a lo largo de todas las situaciones de elección en que se ve envuelto.

3. Inciso 2

Podemos reescribir la ecuación 1.1 para obtener **un modelo de mixed logit** sin una representación de coeficientes aleatorios, sino con una de **componentes de error**.¹ Para ello, podemos descomponer el vector de coeficientes aleatorios en una suma entre su vector de medias poblacionales y un vector de componentes de error. Es decir, dado que $b_n \sim N(\bar{b}, \Sigma)$, entonces podemos decir que $b_n = \bar{b} + v_m$, $v_m \sim N(0, \Sigma)$. Entonces:

$$\begin{aligned} u_{jm} &= \alpha_{jm} + \beta_{mp} price_{jm} + \beta_{mf} feat_{jm} + \varepsilon_{jm} \\ &= b_{jm}^\top z_{jm} + \varepsilon_{jm} \\ &= (\bar{b}_j + v_{jm})^\top z_{jm} + \varepsilon_{jm}, \end{aligned} \quad (3.1)$$

donde $\bar{b}_j = (\bar{\alpha}_j, \bar{\beta}_p, \bar{\beta}_f)^\top$ y $z_{jm} = (1, price_{jm}, feat_{jm})^\top$. De este modo, la **probabilidad de elección incondicional** es

$$\mathbb{P}_{jm} = \int \frac{e^{(\bar{b}_j + v_{jm})^\top z_{jm}}}{\sum_{i \in B} e^{(\bar{b}_i + v_{mi})^\top z_{im}}} f(v_{jm}) dv_{jm}. \quad (3.2)$$

Dada una muestra aleatoria $\left\{ \left(m, s_m, j^*(m) \right) \right\}_{m=1}^M$, la **función de log-verosimilitud** a optimizar es

$$LL(\bar{b}, \Sigma) = \sum_{m=1}^M \sum_{j \in B} y_{jm} \log \mathbb{P}_{jm}; \quad (3.3)$$

sin embargo, esta función de log-verosimilitud es difícil de implementar, dada la integral múltiple contenida en 3.2. Para solucionar esto, podemos condicionar esta probabilidad de elección en v_{jm} , con lo que 3.2 se convierte en una **probabilidad de elección condicional**

$$\mathbb{P}_{jm}^{v_{jm}} = \frac{e^{(\bar{b}_j + v_{jm})^\top z_{jm}}}{\sum_{i \in B} e^{(\bar{b}_i + v_{mi})^\top z_{im}}}. \quad (3.4)$$

Además, en clase vimos que si $\left\{ v_{jm}^r \right\}_{r=1}^R$ son realizaciones del componente de error v_{jm} , entonces

$$\begin{aligned} \mathbb{P}_{jm} &\approx \mathbb{E}_{v_{jm}} [\mathbb{P}_{jm}^{v_{jm}}] \\ &\approx \frac{1}{R} \sum_{r=1}^R \mathbb{P}_{jm}^{v_{jm}^r} \\ &\equiv \hat{\mathbb{P}}_{jm}^R, \end{aligned} \quad (3.5)$$

por lo que podemos reescribir la función de log-verosimilitud usando 3.5 para obtener una **función de log-verosimilitud simulada**:

$$SLL(\bar{b}, \Sigma) = \sum_{m=1}^M \sum_{j \in B} y_{jm} \log \hat{\mathbb{P}}_{jm}^R. \quad (3.6)$$

En esta tarea, optimizaremos 3.6 para obtener estimados de \bar{b}_j y Σ .

¹La única razón siendo que así es como lo practicamos en clase, aunque entiendo que la estimación de un modelo en su forma de coeficientes aleatorios debería ser muy similar.

4. Inciso 3

El siguiente paso permite darle mayor estabilidad a la estimación. Nos apoyaremos del siguiente teorema:

Teorema de la descomposición de Cholesky. *Sea Σ una matriz real, positiva definida, simétrica y de dimensiones $n \times n$. Existe una matriz triangular superior $\Gamma_{n \times n}$ tal que $\Sigma = \Gamma^\top \Gamma$. Además, Γ es única.*

Sea entonces $\omega \sim N(0, \mathbf{I}_{n \times n})$. Como sabemos que Σ es una matriz real de dimensión n , positiva definida y simétrica, entonces por el teorema anterior, $\exists \Gamma$ tal que $\Sigma = \Gamma^\top \Gamma$. Por lo tanto, podemos expresar el componente de error v_{jm} como

$$v_{jm} = \Gamma^\top \omega_{jm}, \quad (4.1)$$

donde Γ es la matriz de Cholesky asociada a Σ . Por tanto, si $\left\{ \omega_{jm}^r \right\}_{r=1}^R$ son realizaciones independientes de ω_{jm} , entonces usamos 4.1 para obtener

$$\hat{\mathbb{P}}_{jm}^R = \frac{1}{R} \sum_{r=1}^R \frac{e^{(\bar{b}_j + \Gamma^\top \omega_{jm})^\top z_{jm}}}{\sum_{i \in B} e^{(\bar{b}_i + \Gamma^\top \omega_{jm})^\top z_{im}}}. \quad (4.2)$$

5. Inciso 4

Nuestro problema de optimización numérica es el siguiente:

$$\begin{aligned} \max_{\bar{b}, \Gamma} \text{SLL}(\bar{b}, \Gamma) &= \sum_{m=1}^M \sum_{j \in B} y_{jm} \log \hat{\mathbb{P}}_{jm}^R, \\ \hat{\mathbb{P}}_{jm}^R &= \frac{1}{R} \sum_{r=1}^R \frac{e^{(\bar{b}_j + \Gamma^\top \omega_{jm})^\top z_{jm}}}{\sum_{i \in B} e^{(\bar{b}_i + \Gamma^\top \omega_{jm})^\top z_{im}}}. \end{aligned}$$

Obteniendo $\hat{\Gamma}_{\text{SMLE}}$ podemos construir $\hat{\Sigma}_{\text{SMLE}}$ usando lo que sabemos de la sección 4.

Primero, definiremos tres vectores de condiciones iniciales. La tabla 1 muestra los valores de dichos vectores: el primero es una opción estándar; la segunda corresponde a los parámetros estimados en la tarea anterior, cuando trabajamos con Nested logit; y la tercera, los del Logit condicional.

Tabla 1: Vectores de condiciones iniciales usados en la estimación del modelo Mixed Logit

CI	1. Estándar	2. Nested logit	3. Logit condicional
$\bar{\alpha}_1$	-1.000	-28.196	-37.058
$\bar{\alpha}_2$	-1.000	0.387	0.487
$\bar{\alpha}_3$	-1.000	1.308	1.388
$\bar{\alpha}_4$	0.000	0.000	0.000
$\bar{\beta}_p$	-1.000	0.734	0.644
$\bar{\beta}_f$	-1.000	-1.930	-3.086
σ_p	0.100	0.100	0.100
σ_f	0.100	0.100	0.100
σ_1	0.100	0.100	0.100
σ_2	0.100	0.100	0.100
σ_3	0.100	0.100	0.100

Aunado a las estimaciones mediante condiciones iniciales distintas, emplearemos tres métodos de estimación: BFGS, L-BFGS-B y Nelder-Mead. La tabla 2 contiene los resultados, ordenados de izquierda a derecha, de menor a mayor según el valor del AIC. Es posible notar que muchos de las estimaciones son muy similares, y que el modelo ganador se repitió bastante.

Tabla 2: Comparación de modelos Mixed Logit: parámetros estimados

Método	BFGS	BFGS	BFGS	L-BFGS-B	L-BFGS-B	L-BFGS-B	Nelder-Mead	Nelder-Mead	Nelder-Mead
Cond. inic.	CI 3	CI 2	CI 1	CI 2	CI 3	CI 1	CI 3	CI 2	CI 1
$\bar{\alpha}_1$	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	16.37
$\bar{\alpha}_2$	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.66	22.60
$\bar{\alpha}_3$	-3.45	-3.45	-3.45	-3.45	-3.46	-3.45	-3.45	-3.54	-66.96
$\bar{\alpha}_4$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\bar{\beta}_p$	-40.47	-40.47	-40.47	-40.47	-40.49	-40.49	-40.46	-40.29	-617.21
$\bar{\beta}_f$	-0.04	-0.04	-0.04	-0.04	-0.05	-0.04	-0.04	-0.06	25.59
σ_p	-2.39	-2.39	-2.39	-2.38	-2.39	-2.40	-2.33	-0.49	-8.65
σ_f	2.41	2.41	2.41	2.41	2.42	2.42	2.41	2.42	-73.48
σ_1	-0.07	-0.07	-0.07	-0.06	-0.07	-0.07	-0.07	0.25	-81.58
σ_2	0.18	0.18	0.18	0.18	0.19	0.19	0.18	-0.15	62.34
σ_3	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.56	-42.23
LL	-2647.49	-2647.49	-2647.49	-2647.49	-2647.49	-2647.49	-2647.49	-2648.00	-2703.46
AIC	5314.98	5314.98	5314.98	5314.98	5314.98	5314.98	5314.98	5315.99	5426.93

Finalmente, la tabla 3 muestra los resultados finales junto con los errores estándar de los parámetros calculados asintóticamente. Además, se presenta una comparación con los modelos de Nested logit y Logit condicional. Es posible notar que el modelo de Mixed logit obtenido arroja un AIC menor, lo que sugiere que es superior al resto.

Tabla 3: Comparación de modelos Logit Condicional, Nested Logit y Mixed Logit

Parámetro	Mixed logit	EE	Nested logit	EE	Logit condicional	EE
β_p	-40.472	3.878	-28.196	3.780	-37.058	2.399
β_f	-0.045	0.256	0.387	0.102	0.487	0.120
α_1	1.483	0.106	1.308	0.079	1.388	0.088
α_2	0.666	0.058	0.734	0.070	0.644	0.054
α_3	-3.454	0.435	-1.930	0.316	-3.086	0.145
α_4	0.000	0.000	0.000	0.000	0.000	0.000
λ_1	—	—	0.702	0.106	—	—
λ_2	—	—	0.525	0.120	—	—
σ_p	-2.387	8.633	—	—	—	—
σ_f	2.414	0.489	—	—	—	—
σ_1	-0.065	0.359	—	—	—	—
σ_2	0.184	0.702	—	—	—	—
σ_3	0.411	0.871	—	—	—	—
AIC	5314.98		5322.20		5327.11	

Nota: Los errores estándar (EE) fueron calculados mediante la teoría asintótica utilizando el hessiano numérico aproximado con `numdifftools`. En el caso del modelo Nested Logit, los parámetros λ_r fueron transformados usando el método delta.