Elección discreta: Teoría, estimación y métodos numéricos
Tarea 3: Nested logit

El Colegio de México
Prof. Edwin Muñoz Rodríguez
Alumno: Max Brando Serna Leyva

Índice

1. Instrucciones iniciales

Vamos a extender la implementación del logit condicional en la Tarea 2 para implementar el Nested Logit usando los datos de la base de datos `yogurt.csv`. Normaliza la utilidad de la alternativa 4, seleccionando $\alpha_4 = 0$. Tu modelo debe tener dos nidos, anida las alternativas del siguiente modo: $\{\{1,2\},\{3,4\}\}$. Permita correlaciones específicas diferentes, esto es $\lambda_1 \neq \lambda_2$. Escribe la expresión analítica para la choice probability y la log-verosimilitud del modelo (asumiendo que se trata de una muestra aleatoria). Trata todas las observaciones como independientes. Esto es un supuesto fuerte *within household*, se corregirá cuando implementemos el latent class model (siguiente tarea).

La **probabilidad de elección de la alternativa** j es

$$\text{Prob}(j \mid s_n; B) = \frac{\left(\sum_{i \in B_r} e^{v(x_i,s_n)/\lambda_r}\right)^{\lambda_r - 1} e^{v(x_j,s_n)/\lambda_r}}{\sum_{l=1}^{R}\left(\sum_{i \in B_l} e^{v(x_i,s_n)/\lambda_l}\right)^{\lambda_l}},$$

y la **función de log-verosimilitud del modelo** es

$$LL\left(\alpha,\beta,\lambda_r \ \mid \{m, s_m, j^*(m)\}_{m=1}^{M}\right) = \sum_{m=1}^{M} \log\left(\frac{\left(\sum_{i \in B_{r^*(m)}} e^{v_{im}/\lambda_{r^*(m)}}\right)^{\lambda_{r^*(m)}-1} e^{v_{j^*(m)m}/\lambda_{r^*(m)}}}{\sum_{l=1}^{R}\left(\sum_{i \in B_l} e^{v_{im}/\lambda_l}\right)^{\lambda_l}}\right),$$

donde

$$v_{im} = \alpha_i + \beta' x_{im},$$
$$v_{j^*(m)m} = \alpha_{j^*(m)} + \beta' x_{j^*(m)m}$$

2. Pregunta 1

Vamos a (intentar) usar chatGPT u otro LLM para extender el código del logit condicional que reportaste en tu Tarea 2 para estimar el Nested Logit. (warning: puede fallar!)

2.1. Inciso a

Usa como input inicial tu código, o partes de tu código, para la estimación del logit condicional, reportado en la Tarea 2.

2.2. Inciso b

Promptea chatGPT de manera que transforme este input inicial en el código para la estimación del Nested Logit. Para ello puede ser muy util pensar en pseudo-código primero. Presta atención a las choice probabilities y a la log-verosimilitud.

2.3. Inciso c

Reporta todos los prompts y respuestas de chatGPT, detente cuando estés conforme con el código que te dio chatGPT o cuando parezca que las cosas no estan yendo a ningún lado. Nota que entre los primeros prompts debe estar tu código para la Tarea 2.

- ¿Estás conforme con el código generado usando GPT?
  - Si si estás conforme, ¿por qué?¿cómo sabes que es correcto?
  - Si no estás conforme, ¿por qué?

Respuesta. Decidí no usar el método propuesto de prompteo en ChatGPT, dado que al intentar dar mi código como input y la fórmula de la log-verosimilitud, el LLM malinterpretaba la fórmula y no captura correctamente algunos detalles. Decidí hacerlo de la misma forma que siempre lo hago, la cual consiste en apoyarse de ChatGPT exclusivamente para algunos detalles como la suma sobre los nidos, o la vectorización de un for loop.

3. Pregunta 2

Si no estás conforme con usar GPT para esto, porque no lleva a ningún lugar, porque toma demasiado tiempo, o porque no confias/entiendes el output de GPT, modifica tu script manualmente. ¡Esto es lo que solíamos hacer antes de 2023!

Respuesta. Decidí implementar la LL de forma manual, usando chatGPT para detalles específicos en el cálculo. Adicionalmente, añadí una condición a la LL en la que $\lambda_r = 1$, para $r = 1, 2$, con la finalidad de comprobar que el modelo colapsa al logit condicional.

4. Pregunta 3

Crea una tabla para comparar los estimados del Nested Logit con los obtenidos usando el logit condicional. Incluye los estimadores de los parámetros y sus errores estándar, así como el Akaike Information Criteria (AIC). Para estimar los errores puede usar método delta o bootstrap, el que prefieras ¿Qué concluyes sobre las correlaciones y los patrones de sustitución a partir de los estimados de λ?

Respuesta. Para la estimación del modelo Nested Logit utilicé cuatro métodos numéricos distintos y cuatro vectores de condiciones iniciales, lo que se traduce en un total de 16 estimaciones. Los métodos de estimación numérica utilizados fueron los de `BFGS, L-BFGS-B, Nelder-Mead` y `Powell`. Los vectores de condiciones iniciales se muestran en la tabla 1. Incluí como cuarto vector los resultados del modelo Logit Condicional de la tarea anterior[1].

Tabla 1: Vectores de condiciones iniciales usados en la estimación

ID	β_1	β_2	α_1	α_2	α_3	α_4	γ_1	γ_2
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	20.000	10.000	-10.000	20.000	10.000	0.000	0.400	0.900
3	5.000	4.183	3.367	2.550	1.733	0.000	0.917	0.100
4	-37.058	0.487	1.388	0.644	-3.086	0.000	-0.500	-0.500

[1] Tomé arbitrariamente valores de $\gamma_r = -0.5$ en este escenario, lo que induce deliberadamente valores bajos de λ. Con esto intento evitar que el algoritmo se vaya rápidamente al caso donde $\lambda = 1$ (Logit Condicional).

La tabla 2 contiene los resultados de las 16 estimaciones usando los cuatro métodos y los cuatro vectores de condiciones iniciales mencionados. Notemos que se encuentran ordenados por el valor del *Akaike Information Criterion* (AIC)[2], y que el mejor modelo se repite en varias ocasiones sin importar el vector de condiciones iniciales y el método de estimación numérica seleccionado.

Tabla 2: Comparación de los 16 modelos Nested Logit ordenados por AIC

Método	**ID**	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	**AIC**
BFGS	1	-28.196	0.387	1.308	0.734	-1.930	0.000	0.702	0.525	5322.20
Nelder-Mead	4	-28.196	0.387	1.308	0.734	-1.930	0.000	0.702	0.525	5322.20
BFGS	3	-28.196	0.387	1.308	0.734	-1.930	0.000	0.702	0.525	5322.20
BFGS	2	-28.196	0.387	1.308	0.734	-1.930	0.000	0.702	0.525	5322.20
BFGS	4	-28.196	0.387	1.308	0.734	-1.930	0.000	0.702	0.525	5322.20
L-BFGS-B	1	-28.196	0.387	1.308	0.734	-1.930	0.000	0.702	0.525	5322.20
L-BFGS-B	4	-28.197	0.387	1.308	0.734	-1.930	0.000	0.702	0.525	5322.20
Powell	4	-28.159	0.405	1.277	0.711	-1.859	0.000	0.715	0.483	5322.62
Nelder-Mead	2	-33.285	0.467	1.317	0.651	-2.334	0.000	0.888	0.654	5324.90
Nelder-Mead	3	-31.258	0.420	1.379	0.741	-2.673	0.000	0.778	0.874	5326.18
L-BFGS-B	3	-35.994	0.484	1.317	0.600	-2.472	0.000	1.000	0.683	5328.35
Powell	1	-35.709	0.496	1.278	0.579	-2.349	0.000	1.000	0.615	5328.86
L-BFGS-B	2	-37.050	0.488	1.387	0.643	-3.082	0.000	1.000	0.998	5331.08
Powell	2	-37.178	0.480	1.390	0.649	-3.080	0.000	1.000	1.000	5331.14
Powell	3	-37.273	0.493	1.402	0.636	-3.084	0.000	1.000	1.000	5331.23
Nelder-Mead	1	-0.724	0.428	0.444	0.572	-0.383	0.000	0.833	0.159	5625.36

Una vez hallado el mejor modelo, decidí calcular los errores estándar de los parámetros de máxima verosimilitud mediante teoría asintótica y técnicas de boostrap. Para el primero de estos casos, debido a que en mi estimación utilicé la siguiente transformación vista en clase para λ_r,

$$\lambda_r = g(\gamma_r) = \frac{e^{\gamma_r}}{1+e^{\gamma_r}},$$

con la cual nos aseguramos de que $\lambda_r \in (0,1], r = 1,2$. Por tanto, de acuerdo con el método delta, sabemos que

$$\mathrm{Var}(g(\gamma_r)) \approx \left(\frac{\mathrm{d}g(\gamma_r)}{\mathrm{d}\gamma_r}\right)^2 \mathrm{Var}(\gamma_r),$$

por lo que

$$\mathrm{se}(g(\gamma_r)) \approx \left(\frac{\mathrm{d}g(\gamma_r)}{\mathrm{d}\gamma_r}\right) \mathrm{se}(\gamma_r),$$

donde $\mathrm{se}(x) \equiv \sqrt{\mathrm{Var}(x)}$. Además,

$$\begin{aligned}
\frac{\mathrm{d}g(\gamma_r)}{\mathrm{d}\gamma_r} &= \frac{(1+e^{\gamma_r})e^{\gamma_r} - e^{\gamma_r}e^{\gamma_r}}{(1+e^{\gamma_r})^2} \\
&= \frac{e^{\gamma_r}}{1+e^{\gamma_r}}\frac{1}{1+e^{\gamma_r}} \\
&= g(\gamma_r)(1-g(\gamma_r)) \\
&= \lambda_r(1-\lambda_r).
\end{aligned}$$

[2]El cálculo de este valor es el siguiente: $\mathrm{AIC} = 2K - 2\ln(LL)$, donde K es el número de parámetros a estimar y LL es la función de log-verosimilitud.

Es decir, el error estándar de λ_r es

$$\mathrm{se}(\lambda_r) = \lambda_r(1 - \lambda_r)\,\mathrm{se}(\gamma_r).$$

Podemos entonces usar λ_r^{MLE} y $\mathrm{se}(\gamma_r^{MLE})$ para calcular fácilmente $\mathrm{se}(\lambda_r^{MLE})$, junto con el resto de los errores estándar de los estimados. La tabla 3 presenta los resultados del modelo de Nested Logit seleccionado, sus errores estándar asintóticos y los de bootstrap. Para la mayoría de los parámetros, los errores estándar son muy cercanos.

Tabla 3: Estimaciones y errores estándar del modelo Nested Logit

Parámetro	**Estimado**	**EE Asintótico**	**EE Bootstrap**
β_1	-28.1960	3.7797	3.3623
β_2	0.3871	0.1021	0.1087
α_1	1.3084	0.0791	0.0799
α_2	0.7344	0.0698	0.0711
α_3	-1.9296	0.3164	0.3580
α_4	0.0000	0.0000	0.0000
λ_1	0.7017	0.1062	0.0989
λ_2	0.5252	0.1202	0.1666

Nota: Los errores estándar (EE) asintóticos se calcularon con el hessiano numérico aproximado utilizando `numdifftools` y el método delta. Los errores estándar bootstrap se obtuvieron con 3211 réplicas.

Finalmente, a tabla 4 presenta una comparación de resultados del modelo Nested Logit y contra los del logit condicional de la tarea anterior. Notemos que según el valor del AIC, el mejor modelo es el de esta tarea, el Nested Logit.

Tabla 4: Comparación de modelos Logit Condicional y Nested Logit

Parámetro	**Logit condicional**	**EE**	**Nested logit**	**EE**
β_1	-37.058	2.399	-28.196	3.780
β_2	0.487	0.120	0.387	0.102
α_1	1.388	0.088	1.308	0.079
α_2	0.644	0.054	0.734	0.070
α_3	-3.086	0.145	-1.930	0.316
α_4	0.000	0.000	0.000	0.000
λ_1	–	–	0.702	0.106
λ_2	–	–	0.525	0.120
AIC	5327.11		5322.20	

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.

5. Hints y recordatorios

Dado que en el Nested Logit no está garantizado que $LL(\beta)$ sea globalmente cóncava, para la estimación prueben al menos tres algoritmos de optimización numérica diferentes y varios vectores de condiciones iniciales (pueden incluir dentro de esos valores iniciales los estimadores obtenidos con el logit condicional, por ejemplo).

6. Discusión: correlación y patrones de sustitución

El modelo de Nested Logit seleccionado arrojó estimaciones de los parámetros distintas de las obtenidas mediante Logit Condicional, y sus parámetros λ_r nos dicen que si

$$\begin{aligned} \lambda_1 = 0.702 &\Rightarrow \text{corr}(\varepsilon_{1n}, \varepsilon_{2n}) \propto (1 - 0.702)^2 = 0.09, \\ \text{y si } \lambda_2 = 0.525 &\Rightarrow \text{corr}(\varepsilon_{3n}, \varepsilon_{4n}) \propto (1 - 0.525)^2 = 0.23, \end{aligned}$$

es decir, la correlación proporcional entre las alternativas del nido 2 (las alternativas 3 y 4) es mayor que la del nido 1. **Esto sugeriría que los patrones de sustitución son más fuertes entre las alternativas 3 y 4 que entre las alternativas 1 y 2**; de hecho en más del doble.

Notemos también que la correlación proporcional del nido 1 es cercana a cero, además de que varias de las estimaciones de la tabla 2 implican correlaciones proporcionales mucho más cercanas a cero para el nido 1 (pues en ocasiones $\lambda_1 > 0.702$). Esto podría sugerir que la anidación de las alternativas 1 y 2 no es muy robusta.

En suma, a pesar de que los patrones de sustitución son más fuertes en el nido 1 que en el nido 2 para todas las estimaciones, **el modelo Nested Logit parece colapsarse al modelo de Logit Condicional dependiendo de las condiciones iniciales usadas para la estimación numérica y el método numérico.** De nuevo, esto es potencialmente sugestivo de la poca robustez en la anidación de las alternativas.