



Universidad Autónoma de Querétaro  
Facultad de Ingeniería  
Maestría en Ingeniería Matemática

DISEÑO ESTADÍSTICO EXPERIMENTAL PARA LA DISCRIMINACIÓN ENTRE  
DOS MODELOS SOBRE EL COMPORTAMIENTO EN LA TOMA DE DECISIÓN

**TESIS**

Que como parte de los requisitos para obtener el grado de  
Maestra en Ciencias en Ingenieria Matematica

**Presenta:**

LMA. Wilfrido Jacobo Paredes García

**Dirigido por:**

Dr. Eduardo Castaño Tostado

Santiago de Querétaro, enero, 2016



Universidad Autónoma de Querétaro  
Facultad de Ingeniería  
Maestría en Ingeniería Matemática

DISEÑO ESTADÍSTICO EXPERIMENTAL PARA LA DISCRIMINACIÓN ENTRE  
DOS MODELOS SOBRE EL COMPORTAMIENTO EN LA TOMA DE DECISIÓN  
**TESIS**

Que como parte de los requisitos para obtener el grado de  
Maestría en Ciencias en Ingeniería Matemática

**Presenta:**

LMA. Wilfrido Jacobo Paredes García

**Dirigido por:**

Dr. Eduardo Castaño Tostado

SINODALES

Dr. Eduardo Castaño Tostado  
Presidente

Dr. Víctor Manuel Armando Aguirre Torres  
Secretario

Dra. Silvia Lorena Amaya Llano  
Vocal

M en C. Sara Silva Hernández  
Suplente

Dr. Roberto Augusto Gómez Loenzo  
Suplente

Dr. Aurelio Domínguez González  
Director de la Facultad

Firma

Firma

Firma

Firma

Firma

Dra. Ma. Guadalupe Flavia Loarca Piña  
Director de Investigación y Postgrado

Centro Universitario  
Querétaro, Qro.  
Enero de 2016  
México

# Resumen

Los modelos matemáticos son una herramienta que ha permitido durante siglos entender fenómenos en Física, Química, Biología, Ingeniería, e incluso, procesos como la decisión de las personas. Sin embargo, como en todas las áreas, los modelos representan una forma de percibir la realidad, percepción que cambia de persona a persona y, con ello genera con ello modelos rivales junto con la pregunta de cuál es el mejor modelo para un caso particular de estudio. El proceso de toma de decisión no es la excepción, existen una gran variedad de modelos que se basan en diferentes hipótesis, sin embargo, ninguno de ellos ha sido más usado que el modelo Logit Multinomial (MNL) el cual parte del supuesto que todas las alternativas son igualmente atractivas para el decisor, supuesto que en ocasiones no obedece al verdadero proceso de toma de decisión y es necesario evitarlo mediante el uso de un modelo rival, el modelo Logit Multinomial Anidado (NMNL).

El presente trabajo presenta una forma de atacar el problema de confusión entre los modelos (el modelo MNL y el modelo NMNL) mediante la creación de diseños estadísticos óptimos que permitan al investigador conocer las mejores configuraciones de los atributos de las elecciones con el fin de detectar cuál de los dos modelos se ajusta de mejor manera a los datos empíricos recabados.

**Palabras clave:** Evolución Diferencial, Diseño Óptimo, Elección de bebidas carbonatadas, Experimentos de Elección Discreta.



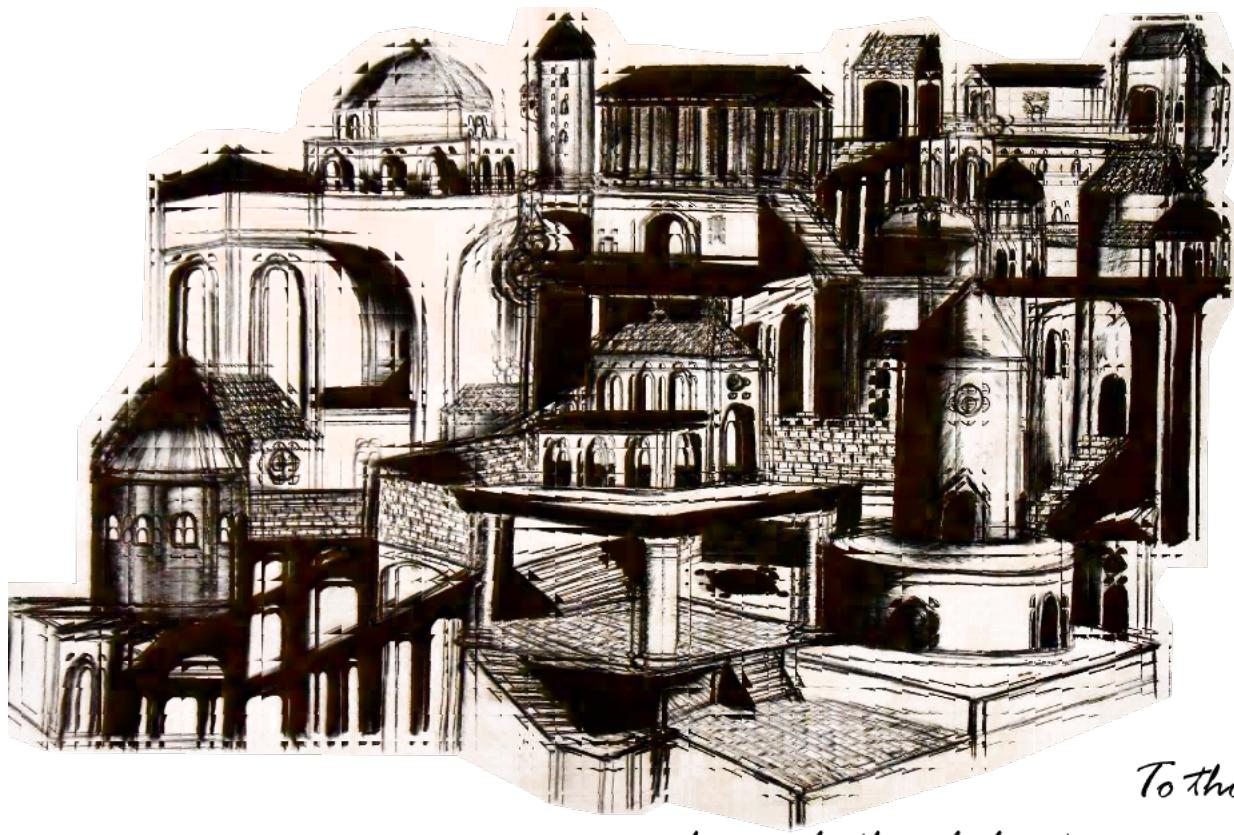
# Summary

Mathematical models are a tool that has allowed for centuries understood phenomena in physics, chemistry, biology, engineering, and even the decision processes of individuals. However, as in all areas, the models represent a way of perceiving reality, perception changes from person to person and therefore it generates with rival models along with the question of what is the best model for a particular case study. The process of decision-making is no exception, there are a variety of models based on different assumptions, however, none has been more used to the multinomial logit model (MNL) which starts from the assumption that all alternatives are equally attractive to the decision maker, of course sometimes does not obey the true decision-making process and you need help using a rival model, the Nested Multinomial Logit (NMNL).

This paper presents a way to attack the problem of confusion between the models (the MNL model and NMNL model) by creating optimal statistical designs that allow the researcher to know the best settings of the attributes of the elections in order to detect Which of the two models better fits the empirical data collected.

**Key words:** Differential Evolution, Optimal Design, Choose of soft drink, Discrete Choice Experiments.





To those  
who made this dedicatory possible



# Agradecimientos

Un agradecimiento total a la Universidad Autónoma de Querétaro y al Consejo Nacional de Ciencia y Tecnología (CONACyT).



# Índice general

<b>Resumen</b>	<b>I</b>
<b>Summary</b>	<b>III</b>
<b>Dedicatoria</b>	<b>v</b>
<b>Agradecimientos</b>	<b>VII</b>
<b>Índice general</b>	<b>XI</b>
<b>1. La complejidad de la toma de decisión</b>	<b>1</b>
1.1. Los retos en el modelado de la toma de decisión . . . . .	1
1.2. La discriminación entre modelos y el diseño experimental . . . . .	4
1.3. El diseño óptimo de experimentos en los DCEs . . . . .	5
1.4. Propuesta y objetivos . . . . .	6
1.4.1. Propuesta . . . . .	6
1.4.2. Objetivos . . . . .	6
<b>I La matriz de información</b>	<b>9</b>
<b>2. Los experimentos de elección discreta</b>	<b>11</b>
2.1. Las preferencias en los DCEs . . . . .	11
2.1.1. Obteniendo las preferencias . . . . .	12
2.1.2. Codificando las respuestas . . . . .	13
2.2. Los modelos en los DCEs . . . . .	13
2.2.1. El modelo logit multinomial . . . . .	13
2.2.2. El modelo logit multinomial anidado . . . . .	15
2.2.3. El significado de discriminar entre el modelo MNL y NMNL . . . . .	16
<b>3. La teoría de diseño óptimo</b>	<b>19</b>
3.1. ¿Qué es un diseño estadístico? . . . . .	19
3.2. Definiendo la matriz de información . . . . .	19
3.2.1. La estimación por mínimos cuadrados . . . . .	20
3.2.2. La estimación por máxima verosimilitud . . . . .	21
3.3. La importancia del diseño experimental . . . . .	22

3.3.1. La imposibilidad de una optimización general . . . . .	22
3.3.2. Normalidad asintótica de $\hat{\theta}_N$ y la matriz de información . . . . .	23
3.3.3. Los criterios de optimalidad . . . . .	24
3.3.4. Propiedades de los criterios de optimalidad . . . . .	26
<b>4. La matriz de información</b>	<b>29</b>
4.1. ¿Qué es un diseño en los DCEs? . . . . .	29
4.2. El trabajo previo de <i>Goos et al.</i> (2010) . . . . .	30
4.3. Un prueba sobre la equivalencia . . . . .	31
4.3.1. El modelo . . . . .	31
4.3.2. Función de log-verosimilitud . . . . .	31
4.3.3. La equivalencia con la codificación $x_{ij} = 0$ . . . . .	33
4.4. El modelo logit multinomial anidado para dos grupos genéricos . . . . .	34
4.4.1. El modelo . . . . .	34
4.4.2. Función de log-verosimilitud . . . . .	34
4.4.3. La matriz de información . . . . .	35
4.4.4. Recuperando el Status Quo con no elección . . . . .	44
<b>II Optimización de la matriz de información</b>	<b>45</b>
<b>5. Algoritmos para la estimación de experimentos</b>	<b>47</b>
5.1. Las condiciones para hallar el óptimo . . . . .	47
5.1.1. Condiciones necesarias y suficientes . . . . .	47
5.1.2. Un objetivo equivalente . . . . .	48
5.2. Un algoritmo de primer orden . . . . .	50
5.2.1. Pasos hacia atrás y hacia adelante . . . . .	50
5.3. La evolución diferencial: una alternativa . . . . .	51
5.3.1. ¿Qué es la evolución diferencial? . . . . .	51
5.3.2. ¿Cómo funciona la evolución diferencial? . . . . .	51
5.3.3. Las ventajas y desventajas de la evolución diferencial . . . . .	52
<b>6. Implementación del algoritmo basado de DE</b>	<b>55</b>
6.1. La función objetivo . . . . .	55
6.2. Especificación del algoritmo . . . . .	56
6.2.1. La codificación de la información . . . . .	56
6.2.2. Los parámetros del algoritmo . . . . .	57
6.3. El algoritmo del intercambio de Fedorov usando DE . . . . .	58
6.4. Construcción de diseños y rendimiento del algoritmo . . . . .	59
<b>7. Diseños óptimos locales para el modelo NMNL</b>	<b>61</b>
7.1. Comparación gráfica . . . . .	61
7.2. Compración vía modelo . . . . .	63
7.3. Más allá de los diseños locales . . . . .	64

<b>III Aplicación</b>	<b>67</b>
<b>8. La elección del refresco</b>	<b>69</b>
8.1. Una breve introducción a la elección del refresco . . . . .	69
8.1.1. El problema . . . . .	69
8.1.2. La elección de un alimento . . . . .	69
8.1.3. Objetivo . . . . .	70
8.2. Metodología . . . . .	70
8.2.1. El modelo . . . . .	70
8.2.2. Las alternativas . . . . .	71
8.2.3. La elección . . . . .	72
8.2.4. Análisis estadístico . . . . .	73
8.3. Resultados y discusión . . . . .	73
8.3.1. Análisis y resultados por grupo: la paradoja de Simpson . . . . .	77
8.4. Conclusión: los marcos de elección . . . . .	78
<b>A. Algoritmos para la obtención de diseños</b>	<b>79</b>
A.1. Algoritmos y funciones base . . . . .	79
A.2. Funciones opcionales . . . . .	84
A.3. Funciones usadas para la comparación . . . . .	84
<b>B. Diseños óptimos locales</b>	<b>87</b>
<b>C. Conjuntos de elección para la decisión del refresco</b>	<b>91</b>



# La complejidad de la toma de decisión

*“Lo sería si lo fuera. Pero si no es no es. Pero si lo fuera podría serlo. Pero si no lo es no lo es y ya”*

— Lewis Carroll, Alicia a través del espejo

## 1.1 Los retos en el modelado de la toma de decisión

El entendimiento del porqué un decisor elige una alternativa sobre otra ha sido siempre un tema de interés; no solo por el entendimiento del proceso psicológico que está involucrado, sino por las consecuencias que conlleva una decisión. El comprender bien esto puede conducir a que un producto sea un éxito o a realizar una inversión en el apoyo de un tratamiento médico que la mayoría de los pacientes desea tomar; mientras que una mala elección puede convertir esas apuestas en un fracaso. La comprensión del comportamiento de los seres humanos frente a una disyuntiva y la complejidad de este proceso radica en que éste es un proceso mental sujeto a *sesgos y marcos* (Tversky y Kahneman, 1974).

El proceso de *toma de decisión* recae en la interacción de la ley de mínimo esfuerzo y dos sistemas mentales, un sistema automático e intuitivo, y un sistema validador y lógico. La conjunción de estos elementos hace que todas las decisiones se vean influenciadas por el sistema intuitivo, y la mayoría de ellas sean tomadas por este mismo. El problema de que las decisiones se basen en la intuición radica en que ésta es altamente influenciable, y por lo tanto, lleva a cometer errores sistemáticos en las decisiones (Kahneman, 2012).

Existen diversos efectos e ilusiones mentales que nos hacen susceptibles a errores sistemáticos y sesgos. Uno de ellos es la facilidad cognitiva, la fluidez con la cual la información viene a la mente, y genera un sesgo en la evaluación de la información debido a que una mayor facilidad cognitiva implica una menor o nula interacción del sistema lógico. Existen varios factores que aumentan la facilidad cognitiva de un mensaje o una idea; la experiencia repetida, es decir, algo parece familiar porque *se cree* que ya se ha experimentado con anterioridad; el “*priming*”, en otras palabras, que un mensaje genera pensamientos e ideas asociados (Berger et al. , 2008); una exposición clara o un mensaje bien escrito y empleo de contrastes y colores por mencionar algunos aunque existen muchos otros factores que generan una facilidad cognitiva. Sin embargo, la facilidad cognitiva no sólo provoca un sesgo a la hora de aceptar una idea, también ayuda a la generación de ideas falsas cuando se combina con los defectos propios (desde el punto de vista racional) del sistema intuitivo. Por ejemplo,

la facilidad cognitiva al estar relacionada disponibilidad de la información se puede emitir un juicio que sobrevalora o infravalora la realidad dependiendo del nivel de disponibilidad esperado (Whittlesea *et al.*, 1990), por ejemplo, en una situación donde alguien ve dos accidentes en una carretera, debido a la disponibilidad de información, podría pensar que es una autopista donde ocurren muchos accidentes, cuando lo más probable es que sean los únicos dos accidentes en autopista que vea en su vida.

La facilidad cognitiva, en muchos casos, es resultado de la conjunción de factores externos con el funcionamiento de nuestro sistema intuitivo, el cual considera que sólo hay lo que se ve. Sin embargo, existen efectos e ilusiones que nos inducen a decisiones con sesgos creados por el propio funcionamiento del sistema intuitivo, el cual funciona por asociatividad de categorías y coherencia (Morewedge y Kahneman, 2010), y como una máquina creadora de conclusiones que está limitada por nuestra capacidad de memoria (Gilbert *et al.*, 1990). El “efecto halo” es uno de los efectos provocados por la asociatividad por categorías, el cual consiste en la caracterización de los atributos de una alternativa; por ejemplo, si una persona cree que una alternativa es buena, entonces pensará que ésta presenta muchos beneficios a bajo costo, mientras que si esta persona piensa que una opción es mala, entonces pensará que ésta presenta pocos beneficios a alto costo. De manera más concreta, si a alguien le agradan las políticas de su presidente, entonces seguramente también le gustará su porte y voz o si alguien no gusta de la forma en que un jugador de soccer patea el balón, entonces ésta creerá que es un mal jugador. El efecto halo también se relaciona con la primera impresión, ésta se puede ejemplificar en el hecho de que si se conoce a alguien inteligente, y después se sabe que también es envidioso, será más aceptado que alguien de quién se supo primero que era envidioso y luego inteligente (Asch, 1946).

Aunado con el funcionamiento del sistema intuitivo, el rechazo del caos y la búsqueda por la causalidad trae consigo la ilusión de confundir plausible con probable (Tversky y Kahneman, 1983). Por ejemplo, en el famoso experimento de Linda, donde se presenta a Linda, una mujer de treinta años, soltera, franca, muy brillante, especializada en filosofía que de estudiante se preocupaba por los asuntos de discriminación y justicia social y que asistió a marchas antinucleares; la mayoría de las personas asignaba una probabilidad mayor al evento de que Linda fuera actualmente una cajera de banco activa en el movimiento feminista que al evento de que Linda fuera simplemente una cajera de banco. Esta búsqueda por la causalidad aumenta la influencia de la “ley de los pequeños números” (Kahneman, 2012), la cual crea la ilusión de la existencia patrones en muestras pequeñas. Por ejemplo, cuando se tienen dos alumnos, uno que obtuvo 10 en sus primeros dos parciales y un 7 en su último parcial y otro que sacó 7 en sus dos primeros parciales y en su último consiguió 10, inmediatamente la mente crea una historia para este hecho como que el primero, alumno notable, se confió porque se suponía aprobado y el segundo, alumno promedio, viéndose en aprietos se esforzó por el 10. La causalidad inventada hace que las personas crean que tienen un alumno sobresaliente y uno dentro de la media cuando en realidad tiene dos estudiantes comunes, los cuales tienen una calificación que tiende a la media.

Por otro lado, el sistema intuitivo tiene la habilidad de estar constantemente alerta del peligro, el cual ha permitido a todo ser vivo y en particular al ser humano aumentar sus probabilidades de supervivencia y reproducción. Sin embargo, la aversión o no al riesgo genera decisiones con sesgo, sesgo hacia la opción segura o hacia el peligro. Por ejemplo, en la disyuntiva de elegir entre ganar 240 pesos o aceptar un trato con 25 % de probabilidad

de ganar 1000 pesos y 75 % de probabilidad de ganar nada, la mayoría de la gente elige la primera opción; mientras que en el dilema de elegir entre perder 750 pesos o jugársela en un trato donde hay 75 % de probabilidad de perder 1000 pesos y 25 % de perder nada, la mayoría de las personas eligen la segunda opción (Rabin *et al.*, 2000).

Los efectos e ilusiones, e incluso la aversión al riesgo, dependen de la situación del decisor o punto de referencia (Tversky y Kahneman, 1991). Hace más de 200 años, Danielle Bernoulli postulaba en su “*Teoría de Utilidad Esperada*” que las personas eligen de acuerdo a la alternativa que más utilidad les proporcione. Sin embargo, esta teoría falla al explicar porqué la gente paga por jugar en sorteos, donde la utilidad esperada de jugar es menor que la utilidad de no jugar; por ejemplo en un sorteo de Melate la utilidad de jugar es de 12.51 pesos y la utilidad esperada de no jugar es de 15.00 pesos; de igual manera falla al dar una explicación del porqué la gente es incapaz de intercambiar bienes por otros de mayor valor o porque la gente se resiste a una pérdida y otras la aceptan sin más, como cuando alguien no vende un boleto para un concierto aun cuando le ofrecen el triple de su valor o cuando dos personas están por perder 100 pesos y una de ellas siente que es una gran pérdida y la otra lo ve con futilidad.

La consideración de que las decisiones son susceptibles a sesgo y que dependen de un punto de referencia está considerada en la “*Teoría de perspectivas*” (Kahneman y Tversky, 1984). Si se deseara saber porqué las personas prefieren una bebida de cafÃ© sobre otra, la teoría sugiere considerar el perfil de publicidad de la bebida, el sabor, el mensaje que crea en el consumidor, etc. Siendo ideal enfocar recursos en todos los rubros benéficos; sin embargo, la limitación de dinero hará que solo se pueda atacar un número limitado de problemas y ello traerá la pregunta sobre cuáles son los atributos con mayor impacto en la elección de una bebida cafÃ©. Una forma de dar respuesta a la pregunta anterior es con el empleo de los *Experimentos de Elección Discreta* (DCEs por sus siglas en inglés). Los DCEs es el nombre general que recibe una de las metodologías existentes para el estudio del comportamiento en la toma decisión. A diferencia de otros métodos para el análisis de la toma de decisión, los DCEs están sustentados en una teoría que lleva varios años de haber sido probada y validada, la “*Teoría de Utilidad Aleatoria*” (Thurstone, 1994). Además de poseer la flexibilidad para representar la aversión al riesgo y la dependencia del punto de referencia (Hardie *et al.*, 1993).

La “*Teoría de Utilidad Aleatoria*” (RUT por sus siglas en inglés) representa una forma de reconocer la incertidumbre que genera el sesgo inherente en las decisiones humanas. La RUT parte del hecho de que toda alternativa es posible de ser elegida y que toda alternativa tiene una probabilidad de ser la más atractiva. El objetivo final es determinar la probabilidad de elección de cada alternativa, en caso de que el decisor elija una de ellas, en función de atributos de interés de las alternativas y determinar cuáles de los atributos que caracterizan a las alternativas tienen un mayor efecto, tanto positivo como negativo, en la probabilidad. Hoy en día existen una gran cantidad de modelos estadísticos basados en la RUT y cada uno de ellos ofrece explicaciones diferentes, las cuales se traducen en la consideración o no de supuestos en el proceso de elección. Sin embargo, en ocasiones, estadísticamente hablando todos los modelos son igual de factibles y por lo tanto sería ideal tener métodos objetivos y confiables que ayudarán a seleccionar alguno para evitar dejar la decisión a nuestra mente sujetada a sesgos.

## 1.2 La discriminación entre modelos y el diseño experimental

La experimentación es una actividad primordial en la mejora y entendimiento de un sistema, por medio de ésta un investigador puede establecer una relación entre un conjunto de variables de entrada y una o varias respuestas. En ese sentido, los DCEs establecen una relación entre la elección de un decisor (tomada de un conjunto de elección) y la magnitud de los atributos que delinean a cada alternativa. A partir de la experimentación se obtiene la estimación de un modelo que permita dar una interpretación satisfactoria de la realidad. Sin embargo, el conocimiento actual ha creado una gran variedad de modelos para explicar una sistema o fenómeno, y esto puede causar confusión a la hora de decidir cuál modelo es mejor para la situación particular que se quiere explicar. Los DCEs no son la excepción, existen diversos modelos que varían de acuerdo a un conjunto de supuestos. El modelo logit multinomial (MNL) fue el primero de estos modelos, el cual afirma que se puede modelar la probabilidad de que un individuo escogido al azar, elija una alternativa de entre un conjunto de alternativas dado un conjunto de atributos que las caracterizan (McFadden, 1974). Sin embargo, el modelo MNL no considera posibles cambios en la elección ni la relación entre las alternativas. En fechas recientes han surgido modelos en la misma dirección que el modelo MNL y que atacan estos problemas. El modelo logit mixto (MXL) propuesto por Revelt y Train (1998), el modelo jerárquico de Bayes (HB) planteado por Allenby y Rossi (1998) y el modelo de clases latentes (LC) formulado por Greene y Hensher (2003) permiten atacar el problema de las revocaciones en las decisiones; mientras que el modelos logit multinomial anidado (NMNL) permite, de manera parcial, tratar el problema que representa la existencia de una relación jerárquica entre las alternativas (Louviere *et al.*, 2000).

En la actualidad existen diversas formas de medir la calidad de un modelo estimado. Por ejemplo, se puede elegir un modelo sobre otro con base en la amplitud de los intervalos de confianza de los parámetros o de la respuesta estimada. Sin embargo, en ocasiones puede suceder que varios modelos resulten plausibles para la realidad que se desea representar. En estos casos, los modelos pueden ser comparados por su error cuadrático medio (MSE), rechazando los modelos con un MSE mayor, o empleando algún criterio de información como el criterio de información de Akaike (AIC) o el criterio de información bayesiano (BIC) que están basados en la verosimilitud del modelo (Konishi y Kitagawa, 2008). Incluso para el caso del modelo NMNL existe una forma de verificar la existencia o no de la jerarquía propuesta usando los parámetros de inclusión del modelo. El problema de estos métodos de selección de modelos radica en que son medidas de calidad relativas para un conjunto de datos dado, es decir, varían de acuerdo al conjunto de datos.

El diseño experimental es la parte de la experimentación que coadyuva en la selección de modelos de manera eficaz y eficiente. Una etapa del diseño experimental consiste en la determinación de los puntos importantes que hay que medir en el experimento para alcanzar un objetivo, como lo puede ser la discriminación entre dos modelos. En el contexto de los DCEs, el diseño experimental se centra en determinar las alternativas que se le presentarán al decisor en el conjunto de elección para alcanzar un objetivo. El diseño de experimentos considera una numerosa serie de objetivos, algunos de estos objetivos están reflejados en los criterios de optimalidad del diseño. Los estadísticos han provisto de una larga serie de criterios, clasificados con una letra del alfabeto, que permiten alcanzar objetivos tales como minimizar la varianza de los parámetros estimados (A, D y E), minimizar la varianza de la

respuesta predicha (G y V) e incluso discriminar entre dos modelos (T) (Atkinson *et al.* , 2007).

La discriminación de modelos, a diferencia de otros objetivos como minimizar la varianza de la respuesta o de los parámetros, depende de una perspectiva y cada perspectiva implica criterios diferentes. Una de éstas y la más conocida fue propuesta por Atkinson en 1975 y supone la comparación entre dos modelos, donde uno de ellos se considera verdadero y el otro se considera como *posiblemente verdadero* (rival), esta situación implica encontrar las condiciones experimentales (puntos) donde los modelos son más desiguales, bajo la configuración de parámetros del modelo rival que hace que ambos modelos sean los más similares posibles. Otro punto de vista parte de considerar ambos modelos como posiblemente verdaderos, la idea es similar, aunque la expresión matemática cambia debido a que los parámetros de ambos modelos se consideran desconocidos. Sin embargo, dado que en muchos casos la discriminación entre modelos puede darse en modelos anidados, la aplicación de estos criterios puede resultar infructuosa y se deben considerar otros enfoques para ayudar en la disyuntiva.

### 1.3 El diseño óptimo de experimentos en los DCEs

Los DCEs vieron su nacimiento con el fin de tratar problemas surgidos en el área económica desde la década de los 70's. Sin embargo, hasta hace muy pocos años la planificación de los DCEs desde el punto de vista estadístico era hecha tomando diseños prefabricados que existían en algunos manuales (handbooks), usando algún software estadístico, como SAS, e incluso se procedía a hacer los diseño “*a mano*” (Louviere y Lancsar, 2009). La razón fue que para la mayoría de los economistas la teoría de diseño óptimo resultaba desconocida. Uno de los problemas de emplear diseños *genéricos* fue expuesto por Louviere (2006), donde se muestra que el empleo de un diseño fraccionado  $2^3$  puede llevar a la estimación de reglas de decisión equivocadas dependiendo de la fracción elegida, haciendo enfasis en que es necesario tener suerte para encontrar la regla de decisión correcta al emplear diseños fraccionados. El problema expuesto por Louviere posee una solución sencilla, el empleo de un diseño factorial completo; aunque inviable en muchos casos debido a que la cantidad de alternativas crece de manera desmedida.

El enfoque del diseño óptimo de experimentos lleva varios años de probar su eficiencia al reducir el número de corridas experimentales y al mismo tiempo ayudar a cumplir alguno de los objetivos del investigador, por ejemplo, en la estimación de varios modelos químicos cinéticos (Atkinson *et al.* , 2007). Sin embargo, como ya se dijo antes, este enfoque ha sido muy poco explorado en el ámbito de los DCEs. Street y Burgess fueron los primeros en ofrecer diseños óptimos para el modelo de Bradley-Terry (un caso particular del modelo MNL cuando se le presenta a un decisor 2 alternativas) y posteriormente para el modelo MNL general. Los trabajos de Street y Burgess comenzaron con resultados para situaciones relativamente sencillas, como el de ofrecer diseños óptimos y cercanos a lo óptimo para el modelo de Bradley-Terry y el MNL cuando las alternativas se caracterizan con atributos con únicamente dos niveles (Street, 2004) y concluyendo con el desarrollo de la matriz de información para el modelo MNL cuando alternativas se caracterizan con atributos con un número de niveles diferentes entre sí (Burgess y Street, 2005). De manera paralela, Ryan y Skatun

(2004) mostraron que la opción de agregar una alternativa de *Status Quo*, la cual resulta una preferencia viable en situaciones como en elección de un tratamiento médico, que de acuerdo con la teoría de perspectivas, donde una persona puede resistirse al cambio de tratamiento, situación que no es realista representarla con el modelo MNL y sugieren el empleo del modelo NMNL en su lugar debido a la existencia una opción de *Status Quo* que significa la opción de no elegir; mientras el resto de las opciones pueden ser catalogadas dentro de la opción elegir.

La falta de un diseño óptimo para el modelo NMNL tardó varios años en ver la luz. Goos *et al.* (2010) mostró el desarrollo la matriz de información de manera general para atacar varios problemas, en su artículo da la expresión general de la matriz de información para el modelo MNL y para el mismo modelo, cuando el investigador tiene información suficiente para traducir su incertidumbre sobre el atractivo de los atributos en un distribución de probabilidad, en otras palabras, una matriz de información que al ser optimizada bajo un criterio puede ofrecer un diseño óptimo tanto para el modelo MNL como para el modelo MIXL (que supone que existe una distribución de probabilidad sobre los atributos). Además da la expresión general de la matriz de información para el modelo NMNL, en caso particular de la jerarquía de elección delineada por la presencia del *Status Quo*.

En el área de diseño óptimo, la matriz de información es la parte medular para la obtención de diseños experimentales. Sin embargo, una parte adicional del trabajo es la caracterización de las regiones óptimas de diseño, por ejemplo, los puntos de diseño para el ajuste de modelos polinomiales de una variable están relacionados con las derivadas de los polinomios de Legendre Atkinson *et al.* (2007). En ese sentido, Jafari (2013) da una caracterización para diseños óptimos bajo el criterio D de manera local para el modelo MNL. Por otra parte, el problema de optimización puede resultar un desafío, en especial, cuando se tienen modelos no lineales, en esta dirección Demirkale *et al.* (2013) establece algunas formas de obtener diseños D-óptimos bajo el empleo de varias herramientas con el objetivo de estimar únicamente los efectos principales; mientras que Kessels *et al.* (2006) propuso un algoritmo adaptativo para la estimación de diseños bayesianos bajo criterios para reducir la varianza en la predicción.

## 1.4 Propuesta y objetivos

### 1.4.1 Propuesta

Se propone el empleo de un diseño experimental que permita la validación de la jerarquía de decisión propuesta por un investigador, con el fin de que éste pueda dar una interpretación clara de la regla de decisión.

### 1.4.2 Objetivos

#### Objetivo general

Construir diseños que satisfagan las condiciones de optimalidad necesarias para la discriminación entre los modelos NMNL y MNL.

## Objetivos particulares

El siguiente trabajo posee los siguientes objetivos:

1. Desarrollar la matriz de información del modelo NMNL.
2. Proponer un algoritmo alternativo de construcción de los diseños estadísticos experimentales.
3. Usar los diseños en una aplicación de interés.



# Parte I

## La matriz de información

Un objetivo implícito presente en la presente obra consiste en cómo representar la información que proviene del fenómeno llamado “*Toma de decisión*” y cómo mejorarla para una meta particular, la cual en este caso es discriminar entre los modelos NMNL y MNL, sin importar el tipo de respuesta que al final se haya recabado. Si el lector se pregunta cómo, la respuesta está en utilizar la **matriz de información**. El cómo es el fin del contenido en esta primera parte.



# 2

## Los experimentos de elección discreta

Piense que se está por sacar al mercado una nueva bebida, un refresco por ejemplo. Si se conociera la preferencia del consumidor, entonces se podría lanzar una bebida con más posibilidades de competir en el mercado. El objetivo final de los DCEs, es ese, determinar la preferencia del consumidor.

### 2.1 Las preferencias en los DCEs

Existe una gran variedad de formas de obtener información sobre las preferencias de una persona, las cuales dependen de la pregunta que se le haga al decisor. Por ejemplo, cualquiera de las siguientes preguntas da cierta información sobre la preferencias de una persona:

1. ¿Cuál alternativa elegiría?
2. ¿Cuáles alternativas le gustan y cuáles no?
3. ¿Qué tanto le gusta una alternativa?
4. ¿Qué orden de preferencia le daría a las opciones presentadas?

Sin embargo, cada una de esas preguntas proporciona distinta información sobre la preferencia de una persona. Por ejemplo, suponga que la preferencia “real” de la población a la cual se quiere dirigir la nueva bebida es

$$\text{Coca – Cola} > \text{Pepsi} > \text{Caballitos} > \text{Peñafiel}$$

es decir, las personas prefieren primero la Coca-Cola, sino hubiera elegirían Pepsi y sino hubiera ninguna de las anteriores preferiría Caballitos y en una última instancia Peñafiel.

Un decisor al responder la primera pregunta lo único que permite establecer es

$$\text{Coca – Cola} > \text{Pepsi} = \text{Caballitos} = \text{Peñafiel}$$

es decir, prefiere Coca-Cola pero no se puede decir más del resto de las alternativas. Por otro lado cuando alguien responde a la segunda pregunta se puede establecer un orden como el siguiente

$$\text{Coca – Cola} = \text{Pepsi} > \text{Caballitos} = \text{Peñafiel}$$

en otras palabras, se sabe que prefiere Coca-Cola y Pepsi son para el decisor mejores alternativas que Caballitos y Peñafiel. Mientras que, cuando se responder la tercera pregunta la jerarquía de preferencia que se podría deducir es

$$\text{Coca} - \text{Cola} > \text{Pepsi} = \text{Caballitos} > \text{Peñafiel}$$

de otro modo, el decisor pudo asignar una calificación mayor a Coca-Cola, menor a Peñafiel y una igual a Pepsi y Caballitos pero entre la de Coca-Cola y Peñafiel. Finalmente, la respuesta que se obtiene de la última pregunta puede conducir a una preferencia como la planteada al principio.

$$\text{Coca} - \text{Cola} > \text{Pepsi} > \text{Caballitos} > \text{Peñafiel}$$

Sea cual sea la pregunta planteada, la RUT resulta útil para establecer un modelo de probabilidad sobre la utilidad de las alternativas, y por ende, sobre las preferencias. Sin embargo, en los DCEs se plantea siempre la primera pregunta. La razón de ello es la calidad práctica de los datos. *Louviere et al.* (2000) habla de que tanto la primera como la segunda pregunta son más propensas a sufrir conmutaciones entre alternativas elegidas y no elegidas, es decir, es posible que una persona que escogió Coca-Cola elija Pepsi en un futuro; mientras que, la tercera y cuarta preguntas, aunque son más ricas, son preguntas con un grado de dificultad mayor a las preguntas 1 y 2, y por tanto su confiabilidad es menor en comparación a éstas.

### 2.1.1 *Obteniendo las preferencias*

En el mundo de los DCEs existen dos caminos a seguir para poder obtener las preferencias de un grupo de interés; estos dos caminos son las preferencias *reveladas* por la acción de elección y las preferencias *declaradas* en una situación.

Las preferencias reveladas consisten en la elección observada en el mercado por parte del consumidor. Por ejemplo, si una persona va y compra en la tienda de conveniencia una Coca-Cola, entonces se sabe que su elección fue Coca-Cola. Si la tienda de conveniencia sólo tuviera 4 marcas en exhibición; Coca-Cola, Pepsi, Caballitos y Peñafiel. Entonces el orden de preferencia de esa persona sería:

$$\text{Coca} - \text{Cola} > \text{Pepsi} = \text{Caballitos} = \text{Peñafiel}$$

Sin embargo, las personas pudieran no tener el mismo *marco* de elección, es decir, otra persona pudo haber comprado Coca-Cola pero la tienda donde la adquirió tenía 5 marcas distintas. La situación anterior plantea un contexto de elección distinto para cada decisor y esto puede llevar a reglas de decisión distintas, debido a que éstas son susceptibles al marco.

Por otro lado, las preferencias declaradas consisten en la elección tomada por un individuo en una situación hipotética planteada por una encuesta, la cual es nombrada bajo el nombre de **conjunto de elección** (Fig. 2.1). De esta manera, al igual que con las preferencias reveladas, es posible establecer una relación

$$\text{Coca} - \text{Cola} > \text{Pepsi} = \text{Caballitos} = \text{Peñafiel}$$

pero, a diferencia de las anteriores, las preferencias declaradas permiten fijar un marco más igualitario de decisión y explorar alternativas no existentes en el mercado. Sin embargo, las preferencias declaradas son 100 % útiles. En general, el tipo de preferencia a explorar depende enteramente del investigador.

ATRIBUTOS	Coca-Cola Life	Pepsi	Caballitos Sabor	Peñafiel Ligero	Status Quo
<b>INFORMACIÓN GENERAL</b>					
Marca	Coca-Cola	Pepsi Co.	AGA	Peñafiel	
Presentación	600 ml	355 ml	600 ml	600 ml	
Sabor	Cola	Cola	Fresa	Limón	
<b>INFORMACIÓN NUTRIMENTAL</b>					
Calorías	300	220	280	90	
Carbohidratos	12	15	18	10	
Azúcares	10	13	17	9	
Sodio	120 mg	90 mg	130 mg	80 mg	
<b>INFORMACIÓN PARTICULAR</b>					
Precio	\$10.00	\$9.00	\$6.00	\$11.00	
Envase	Pet	Lata	Pet	Pet	
Retornable	No	No	No	No	
Pro-Ambiente	Si	Si	Si	Si	
Endulzante	Stevia	Azúcar	Azúcar	Splenda	
<b>ELECCIÓN</b>					
¿Qué bebida elegiría?					

Figura 2.1: Ejemplo de un conjunto de elección para la elección de carrera. Estilo de encuesta siguiendo el ejemplo de Louviere *et al.* (2000). La configuración de alternativas es meramente ilustrativo.

### 2.1.2 Codificando las respuestas

La manera de codificar la preferencia captada en la encuesta es mediante un vector ( $y_i$ ) de longitud  $J$  con entradas 0 y 1, donde  $J$  representa el número de alternativas presentadas en el conjunto de elección,  $i$  representa al  $i$ -ésimo conjunto de elección, los valores 0 corresponden a las alternativas no elegidas y el valor 1 corresponde a la alternativa elegida. Tomando como referencia la Fig. 2.1, si se hubiera elegido Pepsi, entonces

$$y_i = [0, 1, 0, 0, 0]'$$
 (2.1)

note que si se piensa que cada alternativa posee una probabilidad  $p_{ij}$  (la probabilidad de elegir la alternativa  $j$  en el conjunto de elección  $i$ ) y  $\sum_{j=1}^J p_{ij} = 1$ , entonces  $y_i$  es una muestra aleatoria de tamaño uno de una distribución multinomial. De esta manera si se tuviera una muestra de gran tamaño sería posible estimar la probabilidad de cada alternativa mostrada en un conjunto de elección.

## 2.2 Los modelos en los DCEs

Como se mencionó en el Capítulo 1, existe una gran variedad de modelos que son empleados en los DCEs. Sin embargo para el objetivo de discriminar entre el modelo NMNL y el modelo MNL, los dos modelos que son relevantes son estos mismos modelos. Por lo tanto resulta imperativo describirlos y ese es el objetivo de esta sección.

### 2.2.1 El modelo logit multinomial

Como se ha mencionado, los modelos empleados en los DCEs están sustentados en la RUT. En este sentido, la teoría nos dice que la probabilidad de elegir una alternativa es la

probabilidad de que ésta resulte ser la más útil, es decir:

$$p_{ijt} = \Pr_t [j \in i|x] = P[U_{ijt} > U_{ij't}|x] \quad (2.2)$$

donde  $U_{ijt}$  representa la utilidad de la alternativa  $j$  en el conjunto de elección  $i$  para un decisor particular  $t$ . Sin embargo, el proceso de elección es complejo, por lo que es imposible observar la utilidad “completa” por medio de una encuesta. Por lo que se supone que la utilidad de una alternativa dado un conjunto de atributos puede ser vista como:

$$U_{ijt} = x_{ij}\beta + \epsilon_{ijt} \quad (2.3)$$

donde  $x_{ij}$  es la representación de los niveles de los atributos que describen la alternativa  $j$  en el conjunto de elección  $i$ ,  $\beta$  son los coeficientes que describen las preferencias del decisor y  $\epsilon_{ijt}$  son términos de error que depende del decisor. En otras palabras, la utilidad de una alternativa es la suma de lo observado en la encuesta (representado como una combinación lineal entre los atributos y que se asume igual entre individuos) y lo no observado en la encuesta (la idiosincrasia propia decisor).

De esta manera, despejando de (2.3) y sustituyendo en (2.2) se tiene que

$$p_{ijt} = P[(x_{ij}\beta - x_{ij'}\beta) > (\epsilon_{ij't} - \epsilon_{ijt}) | x] \quad (2.4)$$

es decir, la probabilidad de que un sujeto  $t$  elija una alternativa depende de que la diferencia entre la utilidad observada y la diferencia entre la utilidad no observada. Si a la ecuación anterior se le agrega el supuesto de una población de decisores homogénea entonces

$$p_{ij} = P[(x_{ij}\beta - x_{ij'}\beta) > (\epsilon_{ij'} - \epsilon_{ij}) | x] \quad (2.5)$$

### Los supuestos del modelo logit multinomial

La expresión (2.5) depende entonces de la utilidad observada que es conocida y de la idiosincrasia de la población que es desconocida. Al ser  $\epsilon_{ij}$  desconocido, una forma de lidiar con la incertidumbre es suponer una distribución de probabilidad sobre éste.

Históricamente, la distribución se ha supuesto sobre  $\epsilon_{ij}$  es la distribución Weibull (Gumbel o doble exponencial) ([Louviere et al. , 2000](#)) y debido a que posee todos sus momentos finitos positivos y porque el supuesto lleva a una expresión cerrada en caso de que los  $\epsilon_{ij}$  sean IID ([McFadden, 1974](#)). En este sentido, el primer supuesto del modelo MNL parte de esto, y entonces, se asume que  $\epsilon_{ij}$  es independiente e identicamente distribuido (IID) con distribución Weibull. Desde la perspectiva del proceso de elección, el supuesto de IID se traduce en que todas las alternativas son “juzgadas con la misma moneda” (Fig. 2.2).

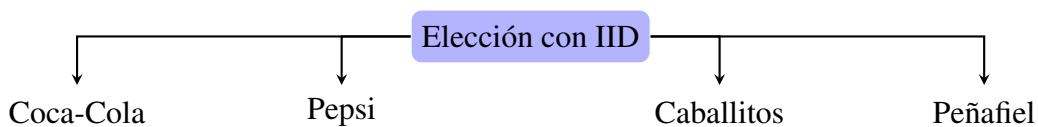


Figura 2.2: Diagrama de elección bajo el supuesto de IID.

Por otro lado, debido a que es imposible exhibir en una encuesta todas las alternativas, un supuesto práctico consistió en asumir que

$$\frac{p_{ij}}{p_{ik}} = \frac{p_{i'j}}{p_{i'k}} \quad (2.6)$$

en otras palabras, todas las alternativas son elegibles, es decir, tienen probabilidad mayor a cero y que una la preferencia de Coca – Cola > Pepsi es independiente de si se le mostró o no al decisor el resto de las alternativas. A la primera parte del supuesto se le conoce como “*supuesto de positividad*”; mientras a la segunda parte se le conoce como “*supuesto de Independencia de las Alternativas Irrelevantes (IIA)*”.

La conjunción de ambos supuestos, permite convertir a la expresión (2.5) en la siguiente expresión

$$p_{ij} = \frac{e^{x'_{ij}\beta}}{\sum_{k=1}^J e^{x'_{ik}\beta}} \quad (2.7)$$

al cual se le conoce como el modelo Logit Multinomial.

### 2.2.2 El modelo logit multinomial anidado

Como ya se dijo en la Sección 1.2 el modelo NMNL surge para modelar, de manera parcial, decisiones donde el supuesto de IID no es adecuado. Una manera en que el supuesto de IID no es adecuado, es cuando las alternativas de elección están correlacionadas por alguna característica común desde el punto de vista del decisor (Fig. 2.3).

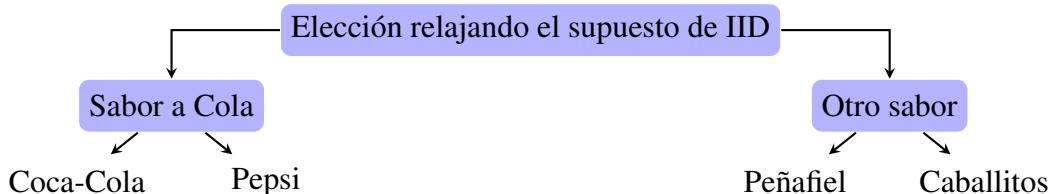


Figura 2.3: Diagrama de elección relajando el supuesto de IID.

Dicho de otra manera, la médula del modelo NMNL consiste en particionar las  $J$  alternativas existentes del conjunto de elección en  $M$  nidos, de esta manera la probabilidad de que una alternativa sea elegida depende de la elección del nido al cual pertenece. Si se supone que la elección de un nido puede ser modelada por el modelo MNL y la elección de las alternativas de la red pueden ser modeladas por el mismo, entonces se tiene un modelo con dos modelos MNL anidados. De esta manera, acorde (2.7), la probabilidad de que un nido  $m$  sea elegido de un conjunto de  $M$  nidos está dada por:

$$p_{i,m} = \frac{e^{\lambda_m V_{i,m}}}{\sum_{k=1}^M e^{\lambda_k V_{i,k}}} \quad (2.8)$$

donde

$$V_{i,m} = \log \left( \sum_{j \in \tau_m} e^{x'_{ij}\beta} \right) \quad (2.9)$$

$V_{i,m}$  es un parámetro de inclusión que mide la utilidad del nido y  $\lambda_m$  es un parámetro de disimilitud que expresa el grado de diferencia entre las alternativas del nido  $m$  (Goos *et al.*, 2010). Mientras que la probabilidad condicional, la probabilidad de elegir una alternativa dentro de la red puede expresarse como

$$p_{ij|m} = \frac{e^{x'_{ij}\beta}}{\sum_{k \in \tau_m} e^{x'_{ik}\beta}} \quad (2.10)$$

donde  $\tau_m$  es un conjunto de indices que contiene únicamente a las alternativas del nido  $m$ .

Por lo tanto, la probabilidad de que una alternativa  $j$  sea elegida dentro del conjunto de elección  $i$  puede expresarse como

$$p_{ij} = p_{ij|m} p_{i,m} \quad (2.11)$$

Note que si  $\lambda_m = \lambda_{m'}$  para todo  $j \neq i$  entonces (2.8) se reduce a

$$\frac{\sum_{j \in \tau_m} e^{x'_{ij}\beta}}{\sum_{k=1}^M \sum_{j \in \tau_k} e^{x'_{ij}\beta}} = \frac{\sum_{j \in \tau_m} e^{x'_{ij}\beta}}{\sum_{k=1}^J e^{x'_{ik}\beta}}$$

y entonces (2.11) es igual a (2.7).

En general los parámetros  $\lambda_m$  (IVs dependiendo del autor) son libres de ser estimados (Louviere *et al.*, 2000), y no son de interés en muchas investigaciones (Goos *et al.*, 2010). Sin embargo, cuando el objetivo de la investigación es la discriminación entre ambos modelos, los parámetros  $\lambda_m$  son los que permiten diferenciar entre ellos y por lo tanto su estimación debe ser considerada y puesta como prioridad.

### 2.2.3 El significado de discriminar entre el modelo MNL y NMNL

Desde el punto de vista de una estructura de decisión el poder ayudar a distinguir entre los modelos MNL y el modelo NMNL se reduce a poder distinguir entre una estructura de decisión y otra, es decir, la capacidad de detectar estadísticamente las diferencias entre las estructuras planteadas en la Fig. 2.2 y la Fig. 2.3.

Pero, ¿qué significa el poder distinguir entre ambas estructuras? El significado desde el punto de vista de la teoría de la que se parte a modelar, radica en que el modelo MNL, al encontrar las características que hacen a una alternativa atractiva útil o no atractiva/inútil son relevantes sin importar de cuál alternativa se trate. Mientras que en el modelo NMNL, la utilidad o inutilidad de una alternativa puede ser sobrevalorada o subestimada dependiendo del grupo al cual pertenezca. De esta manera, la conclusión del modelo MNL se refiere a una configuración óptima; mientras que la conclusión del modelo NMNL se enfoca a una configuración y un grupo al cual pertenecer o no pertenecer. En el caso particular del *Status Quo* (SQ) la distinción entre las jerarquías se traduce en la interpretación que da el decisior a la opción del SQ, es decir, el como ve el decisior al SQ. En algunos casos como en los tratamientos médicos este puede ser parafraseado como “*¿Desea participar?*” Ryan y Skatun (2004) de manera que existe una liga entre la primera decisión de participar y la segunda respuesta en caso de una afirmativa. Sin embargo, esto no significa que el SQ no pueda emplearse bajo el modelo MNL; pero su utilización implica que se supone que el SQ es

considerado una opción del mismo tipo que el resto de las alternativas y hace imposible obtener otro tipo de interpretación de éste.

De este modo, dependiendo de la interpretación las decisiones a tomar pueden resultar más costosas o menos prácticas y este efecto es el que al final justifica la discriminación entre los modelos más allá de la verdad absoluta que exista.



# 3

## La teoría de diseño óptimo

### 3.1 ¿Qué es un diseño estadístico?

Como se mencionó en la Sección 1.2, el fin de la experimentación es establecer la relación entre una respuesta ( $\zeta$ ) y un conjunto de predictores ( $\mathbf{x}$ ), la cual es representada por algún modelo matemático y que puede expresarse como:

$$\zeta = \eta(\mathbf{o}, \theta) + \epsilon \quad (3.1)$$

donde  $\theta = \{\theta_1, \theta_2, \dots, \theta_m\}$  es un conjunto de parámetros desconocidos y  $\epsilon$  representa el error experimental. De manera general, el modelo descrito en (3.1) puede clasificarse en dos grupos respecto de  $\theta$ , *lineal* si  $\eta(\mathbf{o}, \theta) = \theta^T \mathbf{f}(\mathbf{o})$  o *no lineal* en  $\theta$ .

Sin importar el tipo de modelo que el investigador elija, la meta de éste es reunir un número de observaciones ( $\zeta_{n,r}$ ) asociadas a una configuración específica de los predictores ( $\mathbf{o}_n$ ) para poder estimar  $\theta$ . En este sentido, se llama *diseño* a la colección de  $\mathbf{o}_n$ 's tal que

$$\xi_N = \left\{ \begin{array}{ccc} \mathbf{o}_1 & \dots & \mathbf{o}_N \\ \omega_1 & \dots & \omega_N \end{array} \right\} \quad (3.2)$$

donde  $\omega_n = r_n/N$  y  $r_n$  es el número de réplicas en  $\mathbf{o}_n$  con la restricción  $\sum r_n = N$ . Al igual que los modelos, los diseños pueden separarse en dos grupos, los diseños *exactos* (3.2) y los diseños *continuos* en los cuales no existe restricción en el número de observaciones (N), pero con  $\sum \omega_n = 1$  tal que  $\omega_n \in [0, 1]$ .

### 3.2 Definiendo la matriz de información

La estimación de (3.1) es conocida como *estimación parámetrica*, en la cual una parte central es la modelación de  $\theta$  empleando la *matriz de información* del modelo. Sin embargo, la definición de matriz de información no es única, ésta cambia dependiendo de los supuestos, lo cual se traduce en emplear mínimos cuadrados (LSE, por sus siglas en inglés) o máxima verosimilitud (MLE, por sus siglas en inglés). Para distinguir la matriz de información de ambos métodos, a la matriz de información que resulta al emplear el método MLE recibe el nombre de *matriz de información de Fisher*, la cual puede definirse de dos maneras; puede definirse empleando la matriz de dispersión de la primera derivada respecto de los parámetros del logaritmo de la función de densidad del modelo, o como la matriz

esperada de la segunda derivada del logaritmo del modelo (Pukelsheim, 2006). A pesar de que la matriz de información no coincide con la matriz de información de Fisher, ambas comparten dos características muy importantes:

1. Son matrices definidas positivas.
2. Son aditivas, es decir, la matriz asociada a dos puntos es la suma de las matrices asociadas a cada uno de los puntos.

### 3.2.1 La estimación por mínimos cuadrados

El objetivo de la estimación por mínimos cuadrados es encontrar  $\theta$  tal que las diferencias cuadradas entre lo observado y lo que el modelo predice sean mínimas. En ese sentido se tiene

$$\hat{\theta}_N = \arg \min_{\theta} \sum_{n=1}^N r_n [\bar{\zeta}_n - \eta(\mathbf{o}_n, \theta)]^2 \quad (3.3)$$

donde  $\bar{\zeta}_i = \frac{1}{r_n} \sum_{a=1}^{r_n} \zeta_{ij}$ .

#### El caso lineal

Si se considera que el modelo es lineal, entonces (3.3) se puede reescribir como

$$\hat{\theta}_N = \arg \min_{\theta} v(\theta) = \arg \min_{\theta} \sum_{n=1}^N r_n [\bar{\zeta}_n - \theta^T \mathbf{f}(\mathbf{o}_n)]^2 \quad (3.4)$$

dado que el objetivo es la minimización, una forma de alcanzarlo es empleando cálculo diferencial. De esta manera, al derivar  $v(\theta)$  se obtiene

$$\frac{\partial}{\partial \theta} \sum_{n=1}^N r_n [\bar{\zeta}_n - \theta^T \mathbf{f}(\mathbf{o}_n)]^2 = 2 \sum_{n=1}^N r_n \bar{\zeta}_n \mathbf{f}(\mathbf{o}_n) - 2 \sum_{n=1}^N r_n \mathbf{f}(\mathbf{o}_n) \mathbf{f}(\mathbf{o}_n)^T \theta \quad (3.5)$$

Al igualar a cero se obtiene la ecuación “normal” (Fedorov, 2013)

$$\sum_{n=1}^N r_n \mathbf{f}(\mathbf{o}_n) \mathbf{f}(\mathbf{o}_n)^T \theta = \sum_{n=1}^N r_n \bar{\zeta}_n \mathbf{f}(\mathbf{o}_n) \quad (3.6)$$

donde la matriz  $\sum_{n=1}^N r_n \mathbf{f}(\mathbf{o}_n) \mathbf{f}(\mathbf{o}_n)^T$  se puede denotar como  $M_I(\xi_N)$  debido a que depende de  $r_n$  y  $\mathbf{o}_n$ , y es llamada *matriz de información*. La expresión (3.6) define a la matriz de información con base en la ecuación normal, sin embargo esta puede ser definida y obtenida de una segunda manera (aunque ambas son equivalentes). En este sentido, si se reescribe  $v(\theta)$  como

$$\begin{aligned} v(\theta) &= \sum_{n=1}^N r_n [\bar{\zeta}_n - \theta^T \mathbf{f}(\mathbf{o}_n) + \hat{\theta}_N^T \mathbf{f}(\mathbf{o}_n) - \hat{\theta}_N^T \mathbf{f}(\mathbf{o}_n)]^2 \\ &= \sum_{n=1}^N r_n [\bar{\zeta}_n - \hat{\theta}_N^T \mathbf{f}(\mathbf{o}_n)]^2 + \sum_{n=1}^N r_n [(\hat{\theta}_N - \theta)^T \mathbf{f}(\mathbf{o}_n)]^2 \\ &\quad + 2 \sum_{n=1}^N r_n [\bar{\zeta}_n - \hat{\theta}_N^T \mathbf{f}(\mathbf{o}_n)] (\hat{\theta}_N - \theta)^T \mathbf{f}(\mathbf{o}_n) \end{aligned}$$

Se puede obtener, al emplear la igualdad dada por la ecuación normal y la expansión del segundo término, la siguiente expresión

$$v(\theta) = \sum_{n=1}^N r_n [\bar{\zeta}_n - \hat{\theta}_N^T \mathbf{f}(\mathbf{x}_n)]^2 + (\hat{\theta}_N - \theta)^T M(\xi_N) (\hat{\theta}_N - \theta) \quad (3.7)$$

de manera que al derivar la (3.7) dos veces se obtiene

$$\frac{1}{2} \frac{\partial^2 v(\theta)}{\partial \theta \partial \theta^T} = M_I(\xi_N) \quad (3.8)$$

### El caso no lineal en $\theta$

En el caso lineal es posible estimar (3.4) empleando al despejar  $\theta$  de (3.6), siempre y cuando la matriz sea invertible. Sin embargo, cuando  $\eta(\mathbf{o}_n, \theta) \neq \theta^T \mathbf{f}(\mathbf{o}_n)$  entonces no es posible dar una forma cerrada para la estimación de  $\theta$  (Seber y Wild, 2003).

Para tratar el caso no lineal, la idea consiste en aproximar el modelo no lineal usando su expansión por serie de Taylor de manera que

$$\eta(\mathbf{o}, \theta) \approx \eta(\mathbf{o}, \theta_t) + (\theta_t - \theta)^T \mathbf{f}(\mathbf{o}, \theta_t) \quad (3.9)$$

donde  $\mathbf{f}(\mathbf{o}, \theta_t) = \frac{\partial \eta(\mathbf{o}, \theta)}{\partial \theta}$  y  $\theta_t$  es un valor inicial cercano al verdadero valor de  $\theta$ . Así se puede expresar  $v(\bar{\theta})$  como

$$v(\bar{\theta}) = \sum_{n=1}^N r_n [\bar{\zeta}_n^* - \theta^T \mathbf{f}(\mathbf{o}_n)]^2 \quad (3.10)$$

donde  $\bar{\theta} = \theta_t - \theta$  y  $\bar{y}_i^* = \bar{y}_i - \eta(\mathbf{x}, \theta_t)$ , haciendo que la definición del caso lineal de la matriz de información pueda ser empleada en el caso no lineal; aunque ahora depende de  $\theta_t$ .

### 3.2.2 La estimación por máxima verosimilitud

La estimación por máxima verosimilitud tiene como objetivo encontrar el parámetro  $\theta$  que maximice la función de log-verosimilitud, la cual está definida con base en la función de densidad de observaciones ( $p(\bar{\zeta}_{nr} | \mathbf{o}_{nr}, \theta)$ ). De esta manera, se introduce una supuesto sobre la distribución de las observaciones que no se consideró en el método LSE. Además, si éstas son independientes entre sí, se tiene que la función de log-verosimilitud se define como

$$L_N(\theta) = \frac{1}{N} \sum_{n=1}^N \sum_{a=1}^{r_n} \ln p(\bar{\zeta}_{nr} | \mathbf{o}_{nr}, \theta) \quad (3.11)$$

Por otro lado, la función *score* se define como

$$R(\zeta | \mathbf{o}, \theta) = \frac{\partial \ln p(\bar{\zeta}_{nr} | \mathbf{o}_{nr}, \theta)}{\partial \theta} \quad (3.12)$$

De esta manera es posible definir la matriz de información de Fisher en un punto  $\mathbf{o}_i$  como la varianza de la función *score*

$$\mu(\mathbf{o}, \theta) = \text{Var}[R(\zeta|\mathbf{o}, \theta)] = E\left[\frac{\partial \ln p}{\partial \theta} \frac{\partial \ln p}{\partial \theta^T}\right] = -E\left[\frac{\partial^2 \ln p}{\partial \theta \partial \theta^T}\right] \quad (3.13)$$

Así, es posible definir la matriz de información de Fisher para un diseño  $\xi_N$  como:

$$M_F(\xi_N, \theta) = \sum_{n=1}^N r_n \mu(\mathbf{o}_i, \theta) \quad (3.14)$$

de acuerdo a la propiedad aditiva de este tipo de matrices.

### 3.3 La importancia del diseño experimental

#### 3.3.1 La imposibilidad de una optimización general

Los métodos LSE y MLE resuelven el problema de hallar un valor  $\hat{\theta}$  para su modelo de interés. Sin embargo, ambos métodos dependen del experimento  $\xi$ , y por tanto, existen tantos valores de  $\hat{\theta}$  como experimentos  $\xi$ . En este contexto surge la pregunta de cómo elegir el mejor  $\hat{\theta}$ , la cual puede ser resuelta usando la varianza del estimador. En general, el mejor estimador del parámetro del modelo usando un número  $N$  observaciones será el que cumpla que

$$\text{Var}[\hat{\theta}_N] \geq \text{Var}[\hat{\theta}_N^*] \quad \forall \theta_N \in \Theta \quad (3.15)$$

Sin importar el método de estimación empleado, la matriz de información correspondiente ( $M_I$  o  $M_F$ ) está relacionada con la varianza de  $\hat{\theta}_N$ , la cual puede verse como

$$\text{Var}[\hat{\theta}_N] \approx M^{-1}(\xi_N, \theta) \quad (3.16)$$

donde la igualdad sólo se da en el caso lineal (además de que no hay dependencia de  $\theta$  en ese caso). De este modo, surge la pregunta sobre la existencia de un diseño que permita minimizar  $\text{Var}[\hat{\theta}_N]$  sin importar cómo sea  $\theta$ . La respuesta a esta pregunta es no y la demostración a este hecho está dada con un contra ejemplo (ejemplo 1).

**Ejemplo 1** Consideré el modelo

$$\zeta_{nr} = \theta_t^T f(\mathbf{o}_n) + \epsilon_{nr}$$

en el intervalo  $[-1, 1]$  con  $\theta = [\theta_1, \theta_2]^T$ ,  $f(\mathbf{o}_n) = [1, 0]$ ,  $\epsilon_{nr}$  son variables aleatorias independientes con media cero y varianza constante conocida  $\sigma^2$ . Primero se tiene que calcular  $\mu(x)$

$$\mu(\mathbf{o}) = \sigma^2 f(\mathbf{o}) f^T(\mathbf{o}) = \sigma^2 [1, 0]^T [1, 0] = \begin{bmatrix} 1 & 0 \\ 0 & 0^2 \end{bmatrix}$$

y  $M_F(\xi_N)$ . Dado que es un modelo lineal, entonces

$$M_F(\xi_N) = \sigma^{-2} \sum_n^N r_n f(\mathbf{o}_n) f^T(\mathbf{o}_n) = \sum_n^N r_n \begin{bmatrix} 1 & 0_n \\ 0_n & 0_n^2 \end{bmatrix} = \begin{bmatrix} N & \sum_n^N r_n o_n \\ \sum_n^N r_n o_n & \sum_n^N r_n o_n^2 \end{bmatrix}$$

Debido a que  $M_F(\xi_N)$  debe ser una matriz definida positiva, se probará que el problema de optimización solo se alcanza cuando la matriz no es definida positiva. Así, si  $M_F(\xi_N)$  fuera definida positiva, por definición para cualquier  $z = [z_1, z_2]^\top$  se cumple que

$$z^\top M_F(\xi_N) z > 0 \quad (3.17)$$

Entonces, supóngase que existe  $\xi_N^*$  tal que

$$z^\top (M_F(\xi_N^*) - M_F(\xi_N)) z > 0$$

para todo  $\xi_N$ . Desarrollando se tiene

$$\begin{aligned} z^\top (M_F(\xi_N^*) - M_F(\xi_N)) z &= z^\top \left( \begin{bmatrix} N & \sum_n^N r_n^* o_n^* \\ \sum_n^N r_n^* o_n^* & \sum_n^N r_n^* o_n^{*2} \end{bmatrix} - \begin{bmatrix} N & \sum_n^N r_n o_n \\ \sum_n^N r_n o_n & \sum_n^N r_n o_n^2 \end{bmatrix} \right) z \\ &= z^\top \begin{bmatrix} 0 & \sum_n^N r_n^* o_n^* - \sum_n^N r_n o_n \\ \sum_n^N r_n^* o_n^* - \sum_n^N r_n o_n & \sum_n^N r_n^* o_n^{*2} - \sum_n^N r_n o_n^2 \end{bmatrix} z \\ &= \begin{bmatrix} z_2 \left( \sum_n^N r_n^* o_n^* - \sum_n^N r_n o_n \right), z_1 \left( \sum_n^N r_n^* o_n^* - \sum_n^N r_n o_n \right) \\ + \\ z_2 \left( \sum_n^N r_n^* o_n^{*2} - \sum_n^N r_n o_n^2 \right) \end{bmatrix} z \\ &= 2z_1 z_2 \left( \sum_n^N r_n^* o_n^* - \sum_n^N r_n o_n \right) + z_2^2 \left( \sum_n^N r_n^* o_n^{*2} - \sum_n^N r_n o_n^2 \right) \end{aligned}$$

Como se debe cumplir que la ecuación anterior sea positiva, entonces  $\sum_n^N r_n^* o_n^{*2} = \sum_n^N r_n o_n^2$  en el intervalo  $[-1, 1]$ . Por lo tanto

$$o_n^* = \max_{o_n \in [-1, 1]} r_n o_n^2 = \pm 1$$

De este modo, el diseño está dado por  $\{o_1^* = -1, o_2^* = 1, r_1^* + r_2^* = N\}$ . Sin embargo no hay valores de  $r_n^*$  que nos den una matriz definida positiva. Por ejemplo si  $z = [0, 1]$  entonces:

$$z^\top (M_F(\xi_N^*) - M_F(\xi_N)) z = r_2^* - r_1^* - r_2 + r_1 \quad (3.18)$$

Si  $r_i^*$  fueran iguales, con un diseño con  $r_1 - r_2 < 0$  como  $r_1 = \frac{N}{2} - 1$  y  $r_2 = \frac{N}{2} + 1$  la matriz no es definida positiva. Si  $r_2^* > r_1^*$ , con un diseño con  $r_1 - r_2 < 0$  como  $r_1 = r_1^* - 1$  y  $r_2 = r_2^* + 1$  la matriz no es definida positiva.

### 3.3.2 Normalidad asintótica de $\hat{\theta}_N$ y la matriz de información

El ejemplo 1 muestra que no es posible encontrar un experimento en el sentido estricto, ya que es imposible encontrar una matriz de información definida positiva. Para resolver el problema se hará un supuesto sobre  $\hat{\theta}_N$ . Considere que

$$\sqrt{N} (\hat{\theta}_N - \theta_t) \rightarrow N(0, M(\xi_N, \theta_t)) \quad (3.19)$$

La (3.19) implica que la función de densidad de  $\bar{\theta} = \hat{\theta}_N - \theta_t$  puede aproximarse asintóticamente como

$$n_{\bar{\theta}} = \frac{N^{m/2}}{(2\pi)^{m/2} \det^{-1/2} M(\xi_N, \theta_t)} \exp \left[ -\frac{N}{2} t^T M(\xi_N, \theta_t) t \right] \quad (3.20)$$

y de esta manera la concentración de la densidad puede expresarse como

$$\varepsilon_{\bar{\theta}} = \{t \in \mathbb{R}^m | t^T M(\xi_N, \theta_t) t \leq 1\} \quad (3.21)$$

La (3.21) recibe el nombre de elipsoide de concentración asintótico<sup>1</sup> y es la parte fundamental para el cálculo de los intervalos de confianza del estimador (Fedorov, 2013); además de poseer varias propiedades geométricas descritas por Pronzato y Pázman (2013) y que están condensadas en el Lema 1.

**Lema 1** *Sea  $A_{m \times m}$  una matriz definida positiva y sea  $\xi_A = \varepsilon_{\bar{\theta}} = \{t \in \mathbb{R}^m | t^T A t \leq 1\}$ . Entonces:*

1.  $\text{Vol}(\varepsilon_A) = \frac{\pi^{m/2}}{\Gamma(m/2+1)} \det^{-1/2} A$ .
2. Para cualquier vector  $c \in \mathbb{R}^p$  se tiene que  $\max_{t \in \varepsilon_A} (c^T t)^2 = c^T A^{-1} c$ . En particular
3.  $\max_{\|c\|=1} = 1/\lambda_{\min}(A) = R^2(\varepsilon_A)$ , donde  $\lambda_{\min}(A)$  es el eigenvalor mínimo de  $A$  y  $R^2(\varepsilon_A)$  es el radio de la bola más pequeña que contiene a  $\varepsilon_A$ . Además la longitud de los ejes principales de  $\varepsilon_A$  está dada por  $2/\sqrt{\lambda_i(A)}$ .
4. El cuadrado de la longitud de la media diagonal del paralelepípedo que contiene a  $\varepsilon_A$  y la paralelo al eje coordenado del espacio Euclídeo  $\mathbb{R}^p$  está dado por traza ( $A^{-1}$ )
5. Sean  $\varepsilon_A$  y  $\varepsilon_B$  dos elipsoides definidos similarmente salvo por la matriz que las define, entonces es equivalente
  - $\varepsilon_A \subseteq \varepsilon_B$ .
  - $A - B$  es una matriz definida no negativa.
  - Para cualquier vector  $c \in \mathbb{R}^m$  se tiene que  $c^T A^{-1} c \leq c^T B^{-1} c$ .

### 3.3.3 Los criterios de optimalidad

El Lema 1 es el sustento de los criterios de optimalidad, y por lo tanto, tienen que ver con los intervalos de confianza del estimador. Si bien no se puede reducir la varianza del estimador propiamente, se puede reducir en algún sentido el volumen que ocupa la región que contiene al verdadero valor del estimador. En este sentido se definen los criterios de optimalidad conocidos como: criterio D-óptimo, criterio E-óptimo y criterio A-óptimo.

---

<sup>1</sup>Generalmente se le llama solo elipsoide de concentración.

- **Criterio D.** El criterio D tiene como objetivo minimizar el volumen descrito en el primer punto del Lema 1.

$$\xi^* = \arg \min_{\xi} \det [M^{-1}(\xi, \theta)] \quad (3.22)$$

El criterio D, además posee la propiedad de ser invariante a la reparametrización, es decir,  $\det [M^{-1}(\xi, \beta)] = \det [M^{-1}(\xi, \theta)]$  con  $\beta = A\theta$ .

- **Criterio E.** El criterio E está basado en el tercer punto del Lema 1. En este sentido busca reducir el volumen que ocupa el elipsoide al reducir la longitud de su eje mayor.

$$\xi^* = \arg \max_{\xi} \lambda_{\min} (M^{-1}(\xi, \theta)) \quad (3.23)$$

- **Criterio A.** El criterio A tiene su raíz en el cuarto punto del Lema 1, el cual busca reducir el cuadrado de la longitud de la media diagonal del paralelepípedo que contiene a la región, es decir, reducir la suma de la varianza asintótica de cada uno de los componentes del estimador.

$$\xi^* = \arg \min_{\xi} \text{traza} [M^{-1}(\xi, \theta)] \quad (3.24)$$

El objetivo de cada uno de los criterios anteriores puede apreciarse en la Fig. 3.1. En general, cada una de las características a optimizar en cada uno de los criterios es un escalar funcional. De esta manera, el problema de optimización para encontrar experimentos se reduce a la optimización de

$$\xi^* = \arg \min \Psi [M(\xi, \theta)] \quad (3.25)$$

donde  $\Psi$  es un escalar funcional, al cual, en la literatura, se le llama *criterio de optimalidad*.

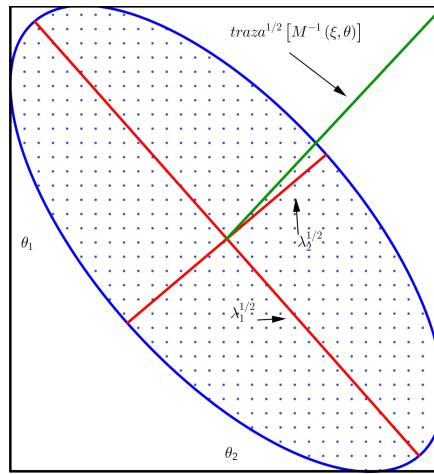


Figura 3.1: Representación del elipsoide de concentración descrito por la (3.21). En azul se aprecia el volumen general que se quiere minimizar al emplear el criterio D; en rojo se ven los dos ejes descritos por los dos eigenvalores de los cuales, el mayor será minimizado en el criterio E; en verde se ve la traza de la matriz de información, la cual está asociada con el paralelepípedo que contiene al elipsoide.

## Un criterio para discriminar entre modelos

*Atkinson et al.* (2007) exponen que, una forma de discriminar entre modelos consiste en emplear diseños que permitan estimar las diferencias entre los modelos, cuando estos son lo más parecidos posibles y suponiendo que se tiene una configuración conocida para alguno de los modelos que se piense como verdadera. En otras palabras

$$\xi^* = \arg \max_{\xi} \int_X [\eta_1(\mathbf{o}) - \eta_2(\mathbf{o}, \hat{\theta})]^2 \xi d\mathbf{o}$$

s.a

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \int_X [\eta_1(\mathbf{o}) - \eta_2(\mathbf{o}, \theta)]^2 \xi d\mathbf{o}$$

donde se supone se conoce una configuración para el modelo  $\eta_1$  y la diferencia radica en la configuración de  $\eta_2$

Sin embargo, esto puede no ser muy conveniente en modelos anidados como lo son el modelo NMNL y el modelo MNL. Ya que una situación óptima, en este caso, es cuando los parámetros que diferencian ambos modelos son nulos. Bajo estos supuestos, cualquier punto es inútil, ya que en realidad los modelos no son parecidos sino iguales.

Entonces, ¿cómo lidiar con la discriminación de modelos anidados? *Dette y Titoff* (2009) proponen suponer que los parámetros que los diferencian no son nulos y aplicar el criterio de *Atkinson et al.* (2007). Sin embargo, la propuesta de esta tesis consiste en encontrar diseños que reduzcan la varianza de los parámetros que hacen diferentes a los modelos, en otras palabras, reducir la varianza de  $\lambda_i$ . De esta manera, si la varianza de tales parámetros es pequeña, los intervalos de confianza de éstos se vuelven pequeños y la seguridad de que éstos sean no nulos aumenta cuando son estimados no nulos y la seguridad de que sean nulos aumenta cuando son estimados nulos. Así, esta seguridad se traduce en un aumento en la capacidad de enfrentar la disyuntiva de cuál modelo es el verdadero, y por lo tanto, los criterios D y A resultan ser una alternativa.

En este sentido, a lo largo de los siguientes capítulos se expondrá el desarrollo de la matriz de información basado en la teoría expuesta en este capítulo para los modelos expuestos en el Capítulo 1.

### 3.3.4 Propiedades de los criterios de optimalidad

Para concluir con este apartado, se mencionarán tres propiedades que deben cumplir todos los criterios de optimalidad y que servirán para poder trazar estrategias para poder dar un diseño. En general, un criterio de optimalidad satisface las propiedades de monotoneidad, homogeneidad y convexidad, las cuales son definidas como sigue

- **Propiedad de monotoneidad.** Para todo criterio  $\Psi$  se cumple que

$$\Psi[M] \geq \Psi[M'] \Rightarrow M' = M + C \quad (3.26)$$

donde  $C$  es una matriz definida positiva, es decir,  $M' \geq M$  bajo el orden de Loewner.

- **Propiedad de homogeneidad.** Para cualquier escalar  $\alpha$  se cumple que

$$\Psi[\alpha M] = \gamma(\alpha) \Psi[M] \quad (3.27)$$

donde  $\gamma$  es una función no creciente.

- **Propiedad de convexidad.** Todo criterio de optimalidad es una función convexa o existe una función equivalente que lo es, y entonces

$$\Psi[M] \leq (1 - \alpha) \Psi[M_1] + \alpha \Psi[M_2] \quad (3.28)$$

donde  $M$  es combinación convexa<sup>2</sup> de  $M_1$  y  $M_2$ .

---

<sup>2</sup> $M$  se dice combinación convexa de  $M_1$  y  $M_2$  si existe  $\alpha \in [0, 1]$  tal que  $M = (1 - \alpha) M_1 + \alpha M_2$ .



# 4

## La matriz de información

Con el contenido expuesto en los capítulos anteriores es posible desarrollar una expresión para la matriz de información del modelo NMNL que nos permita obtener diseños. En este sentido, el objetivo de este capítulo es mostrar un trabajo ya reportado para la matriz de información del modelo NMNL en un caso particular para luego hacer una generalización del mismo que es la propuesta de esta tesis.

### 4.1 ¿Qué es un diseño en los DCEs?

En el caso del modelo NMNL, se tiene que la respuesta  $\zeta_{nr}$  definida en (3.1) está definida como  $\zeta_{nr} = \mathbf{y}_{ir}$ , donde  $\mathbf{y}_i$  está definido como en (2.1) y  $r$  se refiere al  $r$ -ésimo individuo, es decir, las réplicas consisten en entrevistar a otra persona. De esta manera, considerando que la población es homogénea

$$\zeta_{nr} = \mathbf{y}_{ir} = \mathbf{y}_i = [0, \dots, 0, 1, 0, \dots, 0] \quad (4.1)$$

donde cada alternativa tiene una probabilidad  $p_{ij}$  de ser elegida, la cual bajo el modelo MNL se puede ver como  $\frac{e^{x'_{ij}\beta}}{\sum_{k=1}^J e^{x'_{ik}\beta}}$ . Por lo tanto

$$\log(p_{ij}) = \mathbf{x}'_{ij}\beta + \epsilon \quad (4.2)$$

es decir, que el logaritmo de la probabilidad de elegir una alternativa es una combinación lineal de los atributos, lo que hace que el nombre de logit multinomial para el modelo surja de manera natural. Además, noté que la configuración  $(\mathbf{o}_n)$  a la que es hecha cada observación  $\zeta_{n,r} = \mathbf{y}_{i,r}$  es un conjunto de elección. Por lo tanto, la función de log-verosimilitud y la matriz de información tienen como puntos de soporte a conjunto de elección.

Similarmente, bajo el modelo NMNL se tiene que

$$p_{ij} = \frac{e^{x'_{ij}\beta}}{\sum_{k \in \tau_m} e^{x'_{ik}\beta}} \frac{e^{\lambda_m V_{i,m}}}{\sum_{k=1}^M e^{\lambda_k V_{i,k}}} \quad (4.3)$$

donde  $\tau_m$  es el conjunto de índices que contiene a la alternativa  $j$  y  $\lambda_m$  y  $V_{i,m}$  son los parámetros asociados al nido definido por  $\tau_m$ . De manera que el modelo logit queda expresado como

$$\log(p_{ij}) = \mathbf{x}'_{ij}\beta + \lambda_m V_{i,m} - \epsilon - \epsilon_m \quad (4.4)$$

donde  $\epsilon_m$  es el término de error asociado al modelo MNL que modela la elección de los nidos.

## 4.2 El trabajo previo de Goos *et al.* (2010)

Ryan y Skatun (2004) mostraron un ejemplo donde la presencia de una opción de *Status Quo* (SQ), es decir, la opción de no elegir (como el mostrado en la Fig. 2.1) hacia que el supuesto de IID del modelo MNL fuera violado, ya que existía una relación entre todas las alternativas ajenas al SQ: **la decisión de elegir**.

Antes del trabajo de Goos *et al.* (2010) no existían experimentos estadísticos que permitieran mejorar la información que se recababa para el modelo NMNL. La parte central de su trabajo fue ofrecer una expresión para la matriz de información para el modelo NMNL en el caso particular expuesto por Ryan y Skatun (2004), la cual como ya se ha mencionado es la base para estimar experimentos estadísticos.

Ellos inician exponiendo, que para el caso particular donde existen dos grupos; elección (C) y no elección (SQ). De manera que las expresiones 2.8 y 2.9 para este caso se ven como

$$p_{i,C} = \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ} V_{i,SQ}}} \quad (4.5)$$

$$p_{i,SQ} = \frac{e^{\lambda_{SQ} V_{i,SQ}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ} V_{i,SQ}}} \quad (4.6)$$

donde

$$V_{i,C} = \log \left( \sum_{j=1}^J e^{x'_{ij} \beta} \right) \quad (4.7)$$

$$V_{i,SQ} = \log (e^{x_{i,SQ} \beta}) = x_{i,SQ} \beta \quad (4.8)$$

Sin embargo, como la *no elección* solo posee un alternativa, entonces es posible aplicar una simplificación, haciendo que los atributos asociados a la no elección tengan coeficiente 0, de manera que

$$p_{i,C} = \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + 1} \quad (4.9)$$

$$p_{i,SQ} = \frac{1}{e^{\lambda_C V_{i,C}} + 1} \quad (4.10)$$

$$V_{i,SQ} = 0 \quad (4.11)$$

Finalmente, bajo estos supuestos Goos *et al.* (2010) y el hecho de que el modelo puede verse representarse como en (4.4), estiman la función de verosimilitud, la cual posee la siguiente expresión

$$l_i = y_{i,C} V_{i,C} (\lambda_C - 1) - \log (e^{\lambda_C V_{i,C}} + 1) + \sum_{j=1}^J y_{ij} x'_{ij} \beta \quad (4.12)$$

donde J representa las alternativas que no son de no elección,  $y_{ij}$  la respuesta asociada a la alternativa j en el conjunto de elección i y  $y_{i,C} = \sum_{j=1}^J y_{ij}$ . Con base en (4.12) es posible

derivar la matriz de información, la cual queda definida como

$$M = \sum_{i=1}^S M_i = \begin{bmatrix} \mathbf{X}' \mathbf{D} \mathbf{X} & \mathbf{X}' \boldsymbol{\pi} \\ \boldsymbol{\pi}' \mathbf{X} & t \end{bmatrix} \quad (4.13)$$

donde

$$\begin{aligned} \mathbf{X} &= [\mathbf{X}_1' \dots \mathbf{X}_S']' \\ \mathbf{X}_i &= [x_{i1}, \dots, x_{iJ}]' \\ \mathbf{D} &= \text{diag}(\mathbf{D}_1 \dots \mathbf{D}_S) \\ \mathbf{D}_i &= [\mathbf{P}_i + p_{i,C}^{-1} \{ \lambda^2 p_{i,SQ} - 1 \} \mathbf{p}_i \mathbf{p}_i'] \\ \boldsymbol{\pi} &= \lambda [p_{1,SQ} V_{1,C} \mathbf{p}_1', \dots, p_{S,SQ} V_{S,C} \mathbf{p}_S'] \\ t &= \sum_{i=1}^S p_{i,C} p_{i,SQ} V_{i,C}^2 \\ \mathbf{P}_i &= \text{diag}(p_{i1}, \dots, p_{iJ}) \\ \mathbf{p}_i &= [p_{i1}, \dots, p_{iJ}]' \end{aligned}$$

## 4.3 Un prueba sobre la equivalencia

Suponer que los atributos asociados a la no elección tengan un coeficiente 0 como sugiere originalmente, implica que la utilidad de nido asociado al SQ es 0. Sin embargo, ¿qué pasa si se desea partir del supuesto de que la utilidad del SQ no es 0?

### 4.3.1 El modelo

Sea cuál sea la configuración de  $\mathbf{x}_{ij}$ , la utilidad del nido asociado al SQ puede verse como una constante. De esta manera, sin pérdida de generalidad se parte del supuesto de  $\mathbf{x}_{ij} = \mathbf{1}$ . De esta manera las ecuaciones 4.10 y 4.11 se expresan como

$$p_{i,SQ} = \frac{e^{\lambda_{SQ}\beta}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta}} \quad (4.14)$$

$$V_{i,SQ} = \beta \quad (4.15)$$

De esta manera la utilidad de del nido está dado por los atributos asociados a las alternativas.

### 4.3.2 Función de log-verosimilitud

Usando (4.14) y (4.15) para modificar el modelo planteado en la Sección 4.2 se puede desarrollar la función de verosimilitud. Así, se tiene que

$$L_i = \prod p_{ij}^{y_{ij}}$$

Note que se eleva a la  $y_{ij}$  debido a como se definió una observación en (4.3). Así, sustituyendo la (2.11)

$$L_i = \prod (p_{ij|m} p_{i,m})^{y_{ij}}$$

Separando el producto en dos grupos

$$L_i = (p_{ij|SQ} p_{i,N})^{y_{i,SQ}} \prod_{j=1}^J (p_{ij|C} p_{i,C})^{y_{ij}}$$

lo anterior porque solo hay una opción en SQ. Ahora sustituyendo (4.9) y (4.14) se tiene que

$$L_i = \left( p_{ij|SQ} \frac{e^{\lambda_{SQ}\beta}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta}} \right)^{y_{i,SQ}} \prod_{j=1}^J \left( p_{ij|C} \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta}} \right)^{y_{ij}}$$

Note que  $p_{ij|SQ} = 1$  porque solo hay una opción en SQ. Así

$$L_i = \left( \frac{e^{\lambda_{SQ}\beta}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta}} \right)^{y_{i,SQ}} \prod_{j=1}^J \left( p_{ij|C} \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta}} \right)^{y_{ij}}$$

Sustituyendo (2.10)

$$L_i = \left( \frac{e^{\lambda_{SQ}\beta}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta}} \right)^{y_{i,SQ}} \prod_{j=1}^J \left( \frac{e^{x'_{ij}\beta}}{\sum_{k \in \tau_c} e^{x'_{ik}\beta}} \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta}} \right)^{y_{ij}}$$

Note que  $\tau_C = 1 : J$ . Entonces

$$L_i = \left( \frac{e^{\lambda_{SQ}\beta}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta}} \right)^{y_{i,SQ}} \prod_{j=1}^J \left( \frac{e^{x'_{ij}\beta}}{\sum_{k=1}^J e^{x'_{ik}\beta}} \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta}} \right)^{y_{ij}}$$

Sacando los términos que no dependen de  $j$

$$L_i = \frac{e^{y_{i,SQ}\lambda_C V_{i,C}} (e^{\lambda_C V_{i,C}})^{\sum_{j=1}^J y_{ij}}}{(e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta})^{y_{i,SQ}} (e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta})^{\sum_{j=1}^J y_{ij}} \left( \sum_{k=1}^J e^{x'_{ik}\beta} \right)^{\sum_{j=1}^J y_{ij}}} \prod_{j=1}^J e^{y_{ij}x'_{ij}\beta}$$

Sea  $y_{i,C} = \sum_{j=1}^J y_{ij}$  y note que  $y_{i,C} + y_{i,SQ} = 1$ . Entonces

$$L_i = \frac{e^{y_{i,SQ}\lambda_C V_{i,C}} (e^{\lambda_C V_{i,C}})^{y_{i,C}}}{(e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta}) \left( \sum_{k=1}^J e^{x'_{ik}\beta} \right)^{y_{i,C}}} \prod_{j=1}^J e^{y_{ij}x'_{ij}\beta}$$

aplicando logaritmo se tiene

$$\begin{aligned} l_i = & y_{i,SQ}\lambda_C V_{i,C} + y_{i,C}\lambda_C V_{i,C} - \log(e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ}\beta}) \\ & + \\ & -y_{i,C} \log \left( \sum_{k=1}^J e^{x'_{ik}\beta} \right) + \sum_{j=1}^J y_{ij}x'_{ij}\beta \end{aligned}$$

Sustituyendo la (4.7)

$$l_i = y_{i,SQ} \lambda_C V_{i,C} + y_{i,C} \lambda_C V_{i,C} - \log(e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ} \beta}) - y_{i,C} V_{i,C} + \sum_{j=1}^J y_{ij} \mathbf{x}'_{ij} \beta$$

Factorizando  $V_{i,C}$

$$l_i = y_{i,SQ} \lambda_{SQ} \beta + y_{i,C} V_{i,C} (\lambda_C - 1) - \log(e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ} \beta}) + \sum_{j=1}^J y_{ij} \mathbf{x}'_{ij} \beta \quad (4.16)$$

#### 4.3.3 La equivalencia con la codificación $\mathbf{x}_{ij} = \mathbf{0}$

Por un lado se tiene que (4.16), la función de verosimilitud del caso con no elección, define la matriz de información para un conjunto de elección  $i$ , como

$$M_i = E \begin{bmatrix} \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \beta} \right)' & \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' & \left( \frac{\partial l_i}{\partial \lambda_{SQ}} \right) \left( \frac{\partial l_i}{\partial \lambda_{SQ}} \right)' \\ \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \beta} \right)' & \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' & \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_{SQ}} \right)' \\ \left( \frac{\partial l_i}{\partial \lambda_{SQ}} \right) \left( \frac{\partial l_i}{\partial \beta} \right)' & \left( \frac{\partial l_i}{\partial \lambda_{SQ}} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' & \left( \frac{\partial l_i}{\partial \lambda_{SQ}} \right) \left( \frac{\partial l_i}{\partial \lambda_{SQ}} \right)' \end{bmatrix} \quad (4.17)$$

Note que  $\frac{\partial l_i}{\partial \lambda_C}$  en (4.12) y en (4.16) es igual. Además, si  $\left( \frac{\partial l_i}{\partial \beta} \right)_{(4.12)}$  es la derivada parcial de la función de verosimilitud (4.12) y  $\left( \frac{\partial l_i}{\partial \beta} \right)_{(4.16)}$  es la derivada parcial de la función de verosimilitud (4.16), entonces:

$$\left( \frac{\partial l_i}{\partial \beta} \right)_{(4.16)} = y_{i,SQ} \lambda_{SQ} - p_{i,SQ} \lambda_{SQ} + \left( \frac{\partial l_i}{\partial \beta} \right)_{(4.12)} \quad (4.18)$$

y

$$\frac{\partial l_i}{\partial \lambda_{SQ}} = y_{i,SQ} \beta - p_{i,SQ} \beta \quad (4.19)$$

Dado que  $E[y_{i,SQ}] = p_{i,SQ}$ , entonces  $\left( \frac{\partial l_i}{\partial \lambda_{SQ}} \right) \left( \frac{\partial l_i}{\partial \beta} \right)' = \left( \frac{\partial l_i}{\partial \lambda_{SQ}} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' = 0$ . Por lo tanto, se tiene que (4.17) se puede expresar como

$$M_i = E \begin{bmatrix} M'_i & \mathbf{0}_6 \\ \mathbf{0}'_6 & K \end{bmatrix} \quad (4.20)$$

donde  $K$  es una constante positiva determinada por  $\left( \frac{\partial l_i}{\partial \lambda_D} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)'$  y  $M'_i$  es la matriz de información definida en (4.13). De este modo cualquier diseño estadístico óptimo bajo algún criterio, lo será sin importar la codificación de los atributos.

## 4.4 El modelo logit multinomial anidado para dos grupos genéricos

### 4.4.1 El modelo

Para un conjunto de elección  $i$ , se denota la probabilidad de que la alternativa  $j$  sea escogida por  $p_{ij}$ , la probabilidad de que una alternativa sea elegida del grupo C por  $p_{i,C}$  y la probabilidad de que una alternativa sea elegida del grupo D por  $p_{i,D}$ . De esta manera, la probabilidad de que un decisor elija una opción del grupo C es

$$p_{i,C} = \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}}} \quad (4.21)$$

entretanto la probabilidad de que un decisor escoja una opción del grupo D queda modelada como

$$p_{i,D} = \frac{e^{\lambda_D V_{i,D}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}}} \quad (4.22)$$

donde  $\lambda_C$  y  $\lambda_D$  al igual que en el caso particular *Status Quo* son los coeficientes de disimilitud entre el grupo C y el grupo D; mientras que  $V_{i,C}$  y  $V_{i,D}$ , los valores inclusivos de las alternativas dentro de su grupo correspondiente, quedan definidos como

$$V_{i,C} = \log \left( \sum_{j \in \tau_C} e^{x'_{ij} \beta} \right) \quad (4.23)$$

$$V_{i,D} = \log \left( \sum_{j \in \tau_D} e^{x'_{ij} \beta} \right) \quad (4.24)$$

note que se está suponiendo que todas las alternativas son caminos reales de decisión, es decir, no existe la “*No elección*”, y por lo tanto  $\tau_C \cup \tau_D = \{1 : J\}$ .

A diferencia de los casos con *Status Quo* en los cuales existe una alternativa fija que permite la simplificación de las ecuaciones 4.21 y 4.22, el casos de dos grupos genéricos no permite una simplificación inmediata; además que la función de probabilidad  $p_{ij}$  se convierte en una función por casos

$$p_{ij} = \begin{cases} \frac{e^{x'_{ij} \beta}}{\sum_{k \in \tau_C} e^{x'_{ik} \beta}} \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}}} & \text{si } j \in \tau_C \\ \frac{e^{x'_{ij} \beta}}{\sum_{k \in \tau_D} e^{x'_{ik} \beta}} \frac{e^{\lambda_D V_{i,D}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}}} & \text{si } j \in \tau_D \end{cases} \quad (4.25)$$

### 4.4.2 Función de log-verosimilitud

La estimación del modelo, al igual que en los casos anteriores, es por máxima verosimilitud; aunque en este caso particular la función depende de tres parámetros:  $\beta$ ,  $\lambda_C$  y  $\lambda_D$ . Se define  $y_{i,C} = \sum_{j \in \tau_C} y_{ij}$  y  $y_{i,D} = 1 - y_{i,C}$ , en otras palabras,  $y_{i,C}$  y  $y_{i,D}$  indican si la

elección tomada fue elegida pertenece o no al grupo respectivo. De esta manera, la función de verosimilitud puede desarrollarse. Primero considere que

$$L_i = \prod p_{ij}^{y_{ij}}$$

Sustituyendo la (2.11) se tiene

$$L_i = \prod (p_{ij|m} p_{i,m})^{y_{ij}}$$

Se separa el producto en los dos grupos y se consigue la expresión

$$L_i = \prod_{j \in \tau_C} (p_{ij|C} p_{i,C})^{y_{ij}} \prod_{j \in \tau_D} (p_{ij|D} p_{i,D})^{y_{ij}}$$

Sustituyendo la (4.25) se obtiene

$$L_i = \prod_{j \in \tau_C} \left( \frac{e^{x'_{ij}\beta}}{\sum_{k \in \tau_C} e^{x'_{ik}\beta}} \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}}} \right)^{y_{ij}} \prod_{j \in \tau_D} \left( \frac{e^{x'_{ij}\beta}}{\sum_{k \in \tau_D} e^{x'_{ik}\beta}} \frac{e^{\lambda_D V_{i,D}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}}} \right)^{y_{ij}}$$

Separando los términos que no depende de  $j$  se puede reducir a

$$L_i = \begin{cases} \frac{(e^{\lambda_C V_{i,C}})^{\sum_{j \in \tau_C} y_{ij}}}{(\sum_{k \in \tau_C} e^{x'_{ik}\beta})^{\sum_{j \in \tau_C} y_{ij}} (e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}})^{\sum_{j \in \tau_C} y_{ij}}} \prod_{j \in \tau_C} e^{y_{ij} x'_{ij}\beta} \\ \times \\ \frac{(e^{\lambda_D V_{i,D}})^{\sum_{j \in \tau_D} y_{ij}}}{(\sum_{k \in \tau_D} e^{x'_{ik}\beta})^{\sum_{j \in \tau_D} y_{ij}} (e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}})^{\sum_{j \in \tau_D} y_{ij}}} \prod_{j \in \tau_D} e^{y_{ij} x'_{ij}\beta} \end{cases}$$

Agrupando y reduciendo se consigue la expresión

$$L_i = \frac{(e^{\lambda_C V_{i,C}})^{y_{i,C}} (e^{\lambda_D V_{i,D}})^{(1-y_{i,C})}}{(\sum_{k \in \tau_C} e^{x'_{ik}\beta})^{y_{i,C}} (\sum_{k \in \tau_D} e^{x'_{ik}\beta})^{(1-y_{i,C})} (e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}})} \prod_{j=1}^J e^{y_{ij} x'_{ij}\beta}$$

Por lo tanto la función de log-verosimilitud queda determinada por

$$l_i = \begin{cases} y_{i,C} V_{i,C} (\lambda_C - 1) + (1 - y_{i,C}) V_{i,D} (\lambda_D - 1) \\ - \log (e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}}) + \sum_{j=1}^J y_{ij} x'_{ij}\beta \end{cases} \quad (4.26)$$

#### 4.4.3 La matriz de información

Por la definición de matriz de información, ésta puede ser derivada de la función de log-verosimilitud. Entonces para un conjunto de elección particular  $i$  se tiene que la matriz de información del modelo asociado a él puede expresarse como

$$M_i = E \begin{bmatrix} \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \beta} \right)' & \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' & \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' \\ \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \beta} \right)' & \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' & \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' \\ \left( \frac{\partial l_i}{\partial \lambda_D} \right) \left( \frac{\partial l_i}{\partial \beta} \right)' & \left( \frac{\partial l_i}{\partial \lambda_D} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' & \left( \frac{\partial l_i}{\partial \lambda_D} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' \end{bmatrix} \quad (4.27)$$

Debido a la cantidad de cálculos por hacer el desarrollo de la (4.27) se ha dividio en 4 partes.

## Desarrollo de las derivadas parciales

Se tienen tres derivadas parciales por calcular, sin embargo, en todas ellas aparece la función  $V_{i,C}$  y  $V_{i,D}$ . Así, lo primero es calcular las derivadas parciales de estas dos funciones. Usando la (4.25) se tiene

$$\frac{\partial V_{i,C}}{\partial \beta} = \frac{1}{\sum_{j \in \tau_C} e^{x'_{ij}\beta}} \sum_{j \in \tau_C} e^{x'_{ij}\beta} x_{ij} = \sum_{j \in \tau_C} \frac{e^{x'_{ij}\beta}}{\sum_{j \in \tau_C} e^{x'_{ij}\beta}} p_{i,C} p_{i,C}^{-1} x_{ij} = p_{i,C}^{-1} \sum_{j \in \tau_C} p_{ij} x_{ij} \quad (4.28)$$

Análogamente

$$\frac{\partial V_{i,D}}{\partial \beta} = p_{i,D}^{-1} \sum_{j \in \tau_D} p_{ij} x_{ij} \quad (4.29)$$

Con las ecuaciones anteriores ya es posible calcular  $\frac{\partial l_i}{\partial \beta}$ . Así

$$\begin{aligned} \frac{\partial l_i}{\partial \beta} &= y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} \sum_{j \in \tau_C} p_{ij} x_{ij} + (1 - y_{i,C}) (\lambda_D - 1) p_{i,D}^{-1} \sum_{j \in \tau_D} p_{ij} x_{ij} \\ &\quad - \frac{\lambda_C e^{\lambda_C V_{i,C}} p_{i,C}^{-1} \sum_{j \in \tau_C} p_{ij} x_{ij}}{e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}}} - \frac{\lambda_D e^{\lambda_D V_{i,D}} p_{i,D}^{-1} \sum_{j \in \tau_D} p_{ij} x_{ij}}{e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}}} + \sum_{j=1}^J y_{ij} x_{ij} \\ &= y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} \sum_{j \in \tau_C} p_{ij} x_{ij} + (1 - y_{i,C}) (\lambda_D - 1) p_{i,D}^{-1} \sum_{j \in \tau_D} p_{ij} x_{ij} \\ &\quad - \lambda_C \sum_{j \in \tau_C} p_{ij} x_{ij} - \lambda_D \sum_{j \in \tau_D} p_{ij} x_{ij} + \sum_{j=1}^J y_{ij} x_{ij} \end{aligned}$$

Dado que  $V_{i,C}$  y  $V_{i,D}$  no dependen de  $\lambda_C$  y  $\lambda_D$  entonces las derivadas parciales respecto a estas variables son 0. Así

$$\frac{\partial l_i}{\partial \lambda_C} = y_{i,C} V_{i,C} - \frac{e^{\lambda_C V_{i,C}} V_{i,C}}{e^{\lambda_C V_{i,C}} + e^{\lambda_D V_{i,D}}} = y_{i,C} V_{i,C} - p_{i,C} V_{i,C} = V_{i,C} (y_{i,C} - p_{i,C}) \quad (4.30)$$

análogamente:

$$\frac{\partial l_i}{\partial \lambda_D} = V_{i,C} (1 - y_{i,C} - p_{i,D}) \quad (4.31)$$

## Desarrollo de los productos de las fracciones parciales

Con las expresiones de la derivadas parciales es posible desarrollar cada uno de los elementos que componen a la matriz de información ((4.27)).

- $\left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \beta} \right)'$ . El elemento  $\left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \beta} \right)'$  es el producto de dos elementos con cinco sumandos, entonces el resultado de su producto son 25 sumando que se desglosarán uno por uno. Note que  $y_{i,C}^2 = y_{i,C}$  y similarmente  $(1 - y_{i,C})^2 = y_{i,D}^2 = y_{i,D}$ ; además si  $y_{i,C} = 1$  entonces  $1 - y_{i,C} = 0$  y viceversa.

Primer elemento contra los cinco elementos:

$$y_{i,C} (\lambda_C - 1)^2 p_{i,C}^{-2} \left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right) \quad (4.32)$$

$$0 \quad (4.33)$$

$$-y_{i,C} (\lambda_C - 1) \lambda_D p_{i,C}^{-1} \left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right) \quad (4.34)$$

$$-y_{i,C} (\lambda_C - 1) \lambda_D p_{i,C}^{-1} \left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} x'_{ij} \right) \quad (4.35)$$

$$\begin{aligned} & y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} \left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} y_{ij} x'_{ij} \right) \\ & + \end{aligned} \quad (4.36)$$

$$y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} \left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} y_{ij} x'_{ij} \right)$$

Segundo elemento contra los cinco elementos:

$$0 \quad (4.37)$$

$$(1 - y_{i,C}) (\lambda_D - 1)^2 p_{i,D}^{-2} \left( \sum_{j \in \tau_D} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} x'_{ij} \right) \quad (4.38)$$

$$-(1 - y_{i,C}) (\lambda_D - 1) \lambda_C p_{i,D}^{-1} \left( \sum_{j \in \tau_D} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right) \quad (4.39)$$

$$-(1 - y_{i,C}) (\lambda_D - 1) \lambda_D p_{i,D}^{-1} \left( \sum_{j \in \tau_D} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} x'_{ij} \right) \quad (4.40)$$

$$\begin{aligned} & (1 - y_{i,C}) (\lambda_D - 1) p_{i,D}^{-1} \left( \sum_{j \in \tau_D} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} y_{ij} x'_{ij} \right) \\ & + \end{aligned} \quad (4.41)$$

$$(1 - y_{i,C}) (\lambda_D - 1) p_{i,D}^{-1} \left( \sum_{j \in \tau_D} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} y_{ij} x'_{ij} \right)$$

Tercer elemento contra los cinco elementos:

$$-y_{i,C} (\lambda_C - 1) \lambda_C p_{i,C}^{-1} \left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right) \quad (4.42)$$

$$-(1 - y_{i,C}) (\lambda_D - 1) \lambda_C p_{i,D}^{-1} \left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} x'_{ij} \right) \quad (4.43)$$

$$\lambda_C^2 \left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right) \quad (4.44)$$

$$\lambda_C \lambda_D \left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} x'_{ij} \right) \quad (4.45)$$

$$-\lambda_C \left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} y_{ij} x'_{ij} \right) - \lambda_C \left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} y_{ij} x'_{ij} \right) \quad (4.46)$$

Cuarto elemento contra los otros cinco elementos:

$$-y_{i,C} (\lambda_C - 1) \lambda_D p_{i,C}^{-1} \left( \sum_{j \in \tau_D} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right) \quad (4.47)$$

$$-(1 - y_{i,C}) (\lambda_D - 1) \lambda_D p_{i,D}^{-1} \left( \sum_{j \in \tau_D} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} x'_{ij} \right) \quad (4.48)$$

$$\lambda_C \lambda_D \left( \sum_{j \in \tau_D} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right) \quad (4.49)$$

$$\lambda_D^2 \left( \sum_{j \in \tau_D} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} x'_{ij} \right) \quad (4.50)$$

$$-\lambda_D \left( \sum_{j \in \tau_D} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} y_{ij} x'_{ij} \right) - \lambda_D \left( \sum_{j \in \tau_D} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} y_{ij} x'_{ij} \right) \quad (4.51)$$

Quinto elemento contra los otros cinco elementos:

$$y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} \left( \sum_{j \in \tau_C} y_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right) + \quad (4.52)$$

$$y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} \left( \sum_{j \in \tau_D} y_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right) + (1 - y_{i,C}) (\lambda_D - 1) p_{i,D}^{-1} \left( \sum_{j \in \tau_C} y_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} x'_{ij} \right) \quad (4.53)$$

$$(1 - y_{i,C}) (\lambda_D - 1) p_{i,D}^{-1} \left( \sum_{j \in \tau_D} y_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} x'_{ij} \right) + \quad (4.54)$$

$$-\lambda_C \left( \sum_{j \in \tau_C} y_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right) - \lambda_C \left( \sum_{j \in \tau_D} y_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right) \quad (4.55)$$

$$-\lambda_D \left( \sum_{j \in \tau_C} y_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} x'_{ij} \right) - \lambda_D \left( \sum_{j \in \tau_D} y_{ij} x_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} x'_{ij} \right) \quad (4.56)$$

Dado que se tiene que  $E[y_{ij}] = p_{ij}$  se agrupará pensando en ese hecho<sup>1</sup>. Además  $y_{i,A} \sum_{\tau_a} y_{ij} = y_{ij}$  si  $a = A$  y cero en otro caso. Por lo tanto:

---

<sup>1</sup>El lector debe recordar que algunos coeficientes están asociados a  $\left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right)$ , otros a  $\left( \sum_{j \in \tau_C} y_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} x'_{ij} \right)$  y otros a  $\left( \sum_{j \in \tau_C} p_{ij} x_{ij} \right) \left( \sum_{j \in \tau_C} y_{ij} x'_{ij} \right)$ ,

**Coeficientes de**  $(\sum_{j \in \tau_C} p_{ij} x_{ij}) (\sum_{j \in \tau_C} p_{ij} x'_{ij})$

$$y_{i,C} (\lambda_C - 1)^2 p_{i,C}^{-2} \quad (4.57)$$

$$-2y_{i,C} (\lambda_C - 1) \lambda_C p_{i,C}^{-1} \quad (4.58)$$

$$2 (\lambda_C - 1) p_{i,C}^{-1} \quad (4.59)$$

$$\lambda_C^2 \quad (4.60)$$

$$-2\lambda_C \quad (4.61)$$

Las ecuaciones anteriores no se pueden reducir hasta que se calculen las esperanzas, asi que simplemente se tiene

$$y_{i,C} (\lambda_C - 1)^2 p_{i,C}^{-2} - 2y_{i,C} (\lambda_C - 1) \lambda_C p_{i,C}^{-1} + 2 (\lambda_C - 1) p_{i,C}^{-1} + \lambda_C^2 - 2\lambda_C \quad (4.62)$$

**Coeficientes de**  $(\sum_{j \in \tau_D} p_{ij} x_{ij}) (\sum_{j \in \tau_D} p_{ij} x'_{ij})$  Note que los coeficientes son muy similares a los de  $(\sum_{j \in \tau_C} p_{ij} x_{ij}) (\sum_{j \in \tau_C} p_{ij} x'_{ij})$ . Por lo cual se tiene que

$$y_{i,D} (\lambda_D - 1)^2 p_{i,D}^{-2} \quad (4.63)$$

$$-2y_{i,D} (\lambda_D - 1) \lambda_D p_{i,D}^{-1} \quad (4.64)$$

$$2 (\lambda_D - 1) p_{i,D}^{-1} \quad (4.65)$$

$$\lambda_D^2 \quad (4.66)$$

$$-2\lambda_D \quad (4.67)$$

Y de manera análoga las ecuaciones anteriores tampoco se pueden agrupar hasta el cálculo de las esperanzas, por lo cual

$$y_{i,D} (\lambda_D - 1)^2 p_{i,D}^{-2} - 2y_{i,D} (\lambda_D - 1) \lambda_D p_{i,D}^{-1} + 2 (\lambda_D - 1) p_{i,D}^{-1} + \lambda_D^2 - 2\lambda_D \quad (4.68)$$

**Coeficientes de**  $(\sum_{j \in \tau_C} p_{ij} x_{ij}) (\sum_{j \in \tau_D} p_{ij} x'_{ij})$  Los coeficientes son

$$-y_{i,C} (\lambda_C - 1) \lambda_D p_{i,C}^{-1} \quad (4.69)$$

$$y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} \quad (4.70)$$

$$-y_{i,D} (\lambda_D - 1) \lambda_C p_{i,D}^{-1} \quad (4.71)$$

$$y_{i,D} (\lambda_D - 1) p_{i,D}^{-1} \quad (4.72)$$

$$\lambda_C \lambda_D \quad (4.73)$$

$$-\lambda_C \quad (4.74)$$

$$-\lambda_D \quad (4.75)$$

Note que las ecuaciones 4.69 y 4.72 se hacen cero debido al término que los acompaña. De esta manera los coeficientes anteriores se pueden expresar como

$$-y_{i,C} (\lambda_C - 1) \lambda_D p_{i,C}^{-1} - y_{i,D} (\lambda_D - 1) \lambda_C p_{i,D}^{-1} + \lambda_C \lambda_D - \lambda_C - \lambda_D \quad (4.76)$$

**Coefficientes de**  $(\sum_{j \in \tau_D} p_{ij} x_{ij}) (\sum_{j \in \tau_C} p_{ij} x'_{ij})$  Los coeficientes son

$$-y_{i,D} (\lambda_D - 1) \lambda_C p_{i,D}^{-1} \quad (4.77)$$

$$y_{i,D} (\lambda_D - 1) p_{i,D}^{-1} \quad (4.78)$$

$$-y_{i,C} (\lambda_C - 1) \lambda_D p_{i,C}^{-1} \quad (4.79)$$

$$y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} \quad (4.80)$$

$$\lambda_C \lambda_D \quad (4.81)$$

$$-\lambda_C \quad (4.82)$$

$$-\lambda_D \quad (4.83)$$

Similarmente las ecuaciones 4.78 y 4.80 se hacen cero debido al término que los acompaña. Por lo tanto, los coeficientes anteriores se pueden expresar como

$$-y_{i,D} (\lambda_D - 1) \lambda_C p_{i,D}^{-1} - y_{i,C} (\lambda_C - 1) \lambda_D p_{i,C}^{-1} + \lambda_C \lambda_D - \lambda_C - \lambda_D \quad (4.84)$$

Note que el único elemento que no fue agrupado fue el término dado por la (4.56).

- $\left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' y \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)'$

$\left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)'$  se puede expresar como la suma de los siguientes 10 términos

$$y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} V_{i,C} y_{i,C} \sum_{j \in \tau_C} p_{ij} x_{ij} \quad (4.85)$$

$$-y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} V_{i,C} p_{i,C} \sum_{j \in \tau_C} p_{ij} x_{ij} \quad (4.86)$$

$$y_{i,D} (\lambda_D - 1) p_{i,D}^{-1} V_{i,C} y_{i,C} \sum_{j \in \tau_D} p_{ij} x_{ij} \quad (4.87)$$

$$-y_{i,D} (\lambda_D - 1) p_{i,D}^{-1} V_{i,C} p_{i,C} \sum_{j \in \tau_D} p_{ij} x_{ij} \quad (4.88)$$

$$\lambda_C V_{i,C} y_{i,C} \sum_{j \in \tau_C} p_{ij} x_{ij} \quad (4.89)$$

$$-\lambda_C V_{i,C} p_{i,C} \sum_{j \in \tau_C} p_{ij} x_{ij} \quad (4.90)$$

$$\lambda_D V_{i,C} y_{i,C} \sum_{j \in \tau_D} p_{ij} x_{ij} \quad (4.91)$$

$$-\lambda_D V_{i,C} p_{i,C} \sum_{j \in \tau_D} p_{ij} x_{ij} \quad (4.92)$$

$$V_{i,C} y_{i,C} \sum_{j=1}^J y_{ij} x_{ij} \quad (4.93)$$

$$-V_{i,C} p_{i,C} \sum_{j=1}^J y_{ij} x_{ij} \quad (4.94)$$

Note que las ecuaciones 4.85 y 4.86 se pueden reducir y factorizar como

$$(y_{i,C} - y_{i,C} p_{i,C}) (\lambda_C - 1) p_{i,C}^{-1} V_{i,C} \sum_{j \in \tau_C} p_{ij} x_{ij} \quad (4.95)$$

Recuerde que  $y_{i,C} \sum_{j \in \tau_C} y_{ij} x_{ij} = \sum_{j \in \tau_C} y_{ij} x_{ij}$ , de esta manera las ecuaciones 4.93 y 4.94 se pueden expandir y quedar como

$$(1 - p_{i,C}) V_{i,C} \sum_{j \in \tau_C} y_{ij} x_{ij} + V_{i,C} y_{i,C} \sum_{j \in \tau_D} y_{ij} x_{ij} - V_{i,C} p_{i,C} \sum_{j \in \tau_D} y_{ij} x_{ij} \quad (4.96)$$

Finalmente

$$\left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' = 4.95 + 4.87 + 4.88 + 4.89 + 4.90 + 4.91 + 4.92 + 4.96 \quad (4.97)$$

Similarmente  $\left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)'$  se puede expresar como la suma de los siguientes 10 términos:

$$y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} V_{i,D} y_{i,D} \sum_{j \in \tau_C} p_{ij} x_{ij} \quad (4.98)$$

$$-y_{i,C} (\lambda_C - 1) p_{i,C}^{-1} V_{i,D} p_{i,D} \sum_{j \in \tau_C} p_{ij} x_{ij} \quad (4.99)$$

$$y_{i,D} (\lambda_D - 1) p_{i,D}^{-1} V_{i,D} y_{i,D} \sum_{j \in \tau_D} p_{ij} x_{ij} \quad (4.100)$$

$$-y_{i,D} (\lambda_D - 1) p_{i,D}^{-1} V_{i,D} p_{i,D} \sum_{j \in \tau_D} p_{ij} x_{ij} \quad (4.101)$$

$$\lambda_C V_{i,D} y_{i,D} \sum_{j \in \tau_C} p_{ij} x_{ij} \quad (4.102)$$

$$-\lambda_C V_{i,D} p_{i,D} \sum_{j \in \tau_C} p_{ij} x_{ij} \quad (4.103)$$

$$\lambda_D V_{i,D} y_{i,D} \sum_{j \in \tau_D} p_{ij} x_{ij} \quad (4.104)$$

$$-\lambda_D V_{i,D} p_{i,D} \sum_{j \in \tau_D} p_{ij} x_{ij} \quad (4.105)$$

$$V_{i,D} y_{i,D} \sum_{j=1}^J y_{ij} x_{ij} \quad (4.106)$$

$$-V_{i,D} p_{i,D} \sum_{j=1}^J y_{ij} x_{ij} \quad (4.107)$$

Así las ecuaciones 4.100 y 4.101 se pueden reducir y factorizar como

$$(y_{i,D} - y_{i,D} p_{i,D}) (\lambda_D - 1) p_{i,D}^{-1} V_{i,D} \sum_{j \in \tau_D} p_{ij} x_{ij} \quad (4.108)$$

Recuerde que  $y_{i,D} \sum_{j \in \tau_D} y_{ij} x_{ij} = \sum_{j \in \tau_D} y_{ij} x_{ij}$ , de esta manera las ecuaciones 4.106 y 4.107 se pueden expandir como

$$(1 - p_{i,D}) V_{i,D} \sum_{j \in \tau_D} y_{ij} x_{ij} + V_{i,D} y_{i,D} \sum_{j \in \tau_C} y_{ij} x_{ij} - V_{i,D} p_{i,D} \sum_{j \in \tau_C} y_{ij} x_{ij} \quad (4.109)$$

Por lo tanto

$$\left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' = 4.98 + 4.99 + 4.108 + 4.102 + 4.103 + 4.104 + 4.105 + 4.109 \quad (4.110)$$

- $\left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' y \left( \frac{\partial l_i}{\partial \lambda_D} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)'$ . Para el desarrollo de este producto note que

$$\left( \frac{\partial l_i}{\partial \lambda_C} \right)' = \left( \frac{\partial l_i}{\partial \lambda_C} \right) \quad y \quad \left( \frac{\partial l_i}{\partial \lambda_D} \right)' = \left( \frac{\partial l_i}{\partial \lambda_D} \right) \quad (4.111)$$

Al hacer el producto correspondiente se obtiene

$$\left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' = V_{i,C}^2 (y_{i,C} - p_{i,C})^2 = V_{i,C} (y_{i,C} - 2y_{i,C}p_{i,C} + p_{i,C}^2) \quad (4.112)$$

Análogamente

$$\left( \frac{\partial l_i}{\partial \lambda_D} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' = V_{i,D} (y_{i,D} - 2y_{i,D}p_{i,D} + p_{i,D}^2) \quad (4.113)$$

- $\left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' y \left( \frac{\partial l_i}{\partial \lambda_D} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)'$ . El desarrollo de este producto es más sencillo, ya que al multiplicar la (4.30) con la (4.31) se tiene que

$$\begin{aligned} \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' &= \left( \frac{\partial l_i}{\partial \lambda_D} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' \\ &= V_{i,C} V_{i,D} (y_{i,C} - p_{i,C}) (1 - y_{i,C} - p_{i,D}) \end{aligned} \quad (4.114)$$

### Desarrollo de las esperanzas

El cálculos de las esperanzas entonces se deber hacer sobre cada uno de los productos de las derivadas parciales desarrollados en el apartado anterior. Para ello, se tiene que  $p_{i,C} = 1 - p_{i,D}$ ,  $E[y_{ij}] = p_{ij}$ ,  $E[y_{i,C}] = p_{i,C}$  y  $E[y_{i,D}] = p_{i,D}$ .

- $E \left[ \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \beta} \right)' \right]$ . La esperanza de este elemento se divide en la suma de las esperanzas individuales de los términos que tienen asociados los coeficientes descritos por las ecuaciones 4.56, 4.62, 4.76, 4.84 y 4.68. De esta manera se tiene

$$\mathbb{E} [(4.56)]$$

$$\mathbb{E} \left[ \left( \sum_{j=1}^J y_{ij} \mathbf{x}_{ij} \right) \left( \sum_{j=1}^J y_{ij} \mathbf{x}'_{ij} \right) \right] = \mathbb{E} \left[ \sum_{j=1}^J y_{ij} \mathbf{x}_{ij} \mathbf{x}'_{ij} \right] \quad (4.115)$$

$$= \sum_{j=1}^J p_{ij} \mathbf{x}_{ij} \mathbf{x}'_{ij} \quad (4.116)$$

$$\mathbb{E} [(4.62) (\sum_{j \in \tau_C} p_{ij} \mathbf{x}_{ij}) (\sum_{j \in \tau_C} p_{ij} \mathbf{x}'_{ij})]$$

$$[p_{i,C}^{-1} (\lambda_C^2 (1 - p_{i,C}) - 1)] \left( \sum_{j \in \tau_C} p_{ij} \mathbf{x}_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} \mathbf{x}'_{ij} \right) \quad (4.117)$$

$$\mathbb{E} [(4.68) (\sum_{j \in \tau_D} p_{ij} \mathbf{x}_{ij}) (\sum_{j \in \tau_D} p_{ij} \mathbf{x}'_{ij})]$$

$$[p_{i,D}^{-1} (\lambda_D^2 (1 - p_{i,D}) - 1)] \left( \sum_{j \in \tau_D} p_{ij} \mathbf{x}_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} \mathbf{x}'_{ij} \right) \quad (4.118)$$

$$\mathbb{E} [(4.76) (\sum_{j \in \tau_C} p_{ij} \mathbf{x}_{ij}) (\sum_{j \in \tau_D} p_{ij} \mathbf{x}'_{ij})]$$

$$- [\lambda_C \lambda_D] \left( \sum_{j \in \tau_C} p_{ij} \mathbf{x}_{ij} \right) \left( \sum_{j \in \tau_D} p_{ij} \mathbf{x}'_{ij} \right) \quad (4.119)$$

$$\mathbb{E} [(4.84) (\sum_{j \in \tau_D} p_{ij} \mathbf{x}_{ij}) (\sum_{j \in \tau_C} p_{ij} \mathbf{x}'_{ij})]$$

$$- [\lambda_C \lambda_D] \left( \sum_{j \in \tau_D} p_{ij} \mathbf{x}_{ij} \right) \left( \sum_{j \in \tau_C} p_{ij} \mathbf{x}'_{ij} \right) \quad (4.120)$$

Por lo tanto se obtiene la expresión

$$\mathbb{E} \left[ \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' \right] = 4.116 + 4.117 + 4.118 + 4.119 + 4.120 \quad (4.121)$$

- $\mathbb{E} \left[ \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' \right]$ . Note la mayoría de los elementos de la (4.97) al calcular su esperanza se hacen 0, con excepción de la esperanza de los términos de datos por las ecuaciones 4.95 y 4.96. De esta manera se tiene

$$\mathbb{E} \left[ \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' \right] = \frac{(p_{i,C} - p_{i,C}^2) (\lambda_C - 1) p_{i,C}^{-1} V_{i,C} \sum_{j \in \tau_C} p_{ij} \mathbf{x}_{ij}}{(1 - p_{i,C}) V_{i,C} \sum_{j \in \tau_C} y_{ij} \mathbf{x}_{ij}} \quad (4.122)$$

Pero la (4.122) se puede simplificar a:

$$\mathbb{E} \left[ \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' \right] = (1 - p_{i,C}) \lambda_C V_{i,C} \sum_{j \in \tau_C} p_{ij} \mathbf{x}_{ij} \quad (4.123)$$

- $E \left[ \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' \right]$ . Este elemento es análogo al caso anterior, por lo cual

$$E \left[ \left( \frac{\partial l_i}{\partial \beta} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' \right] = (1 - p_{i,D}) \lambda_D V_{i,D} \sum_{j \in \tau_D} p_{ij} x_{ij} \quad (4.124)$$

- $E \left[ \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' \right]$ ,  $E \left[ \left( \frac{\partial l_i}{\partial \lambda_D} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' \right]$  y  $E \left[ \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' \right]$ . Esta esperanzas resultan más sencillas de estimar, para estos casos se tiene que

$$E \left[ \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' \right] = (p_{i,C} - p_{i,C}^2) V_{i,C}^2 = p_{i,C} p_{i,D} V_{i,C}^2 \quad (4.125)$$

$$E \left[ \left( \frac{\partial l_i}{\partial \lambda_D} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' \right] = (p_{i,D} - p_{i,D}^2) V_{i,D}^2 = p_{i,C} p_{i,D} V_{i,D}^2 \quad (4.126)$$

$$E \left[ \left( \frac{\partial l_i}{\partial \lambda_C} \right) \left( \frac{\partial l_i}{\partial \lambda_D} \right)' \right] = E \left[ \left( \frac{\partial l_i}{\partial \lambda_D} \right) \left( \frac{\partial l_i}{\partial \lambda_C} \right)' \right] = 0 \quad (4.127)$$

Por las ecuaciones 4.121, 4.123, 4.124, 4.125, 4.126 y 4.127 se tiene que la matriz de información del modelo NMNL para dos grupos géricos disjuntos puede verse como

$$\begin{bmatrix} 4.121 & 4.123 & (1 - p_{i,D}) \lambda_D V_{i,D} \sum_{j \in \tau_D} p_{ij} x_{ij} \\ 4.123' & p_{i,C} p_{i,D} V_{i,C}^2 & 0 \\ 4.124' & 0 & p_{i,C} p_{i,D} V_{i,D}^2 \end{bmatrix} \quad (4.128)$$

#### 4.4.4 Recuperando el Status Quo con no elección

La matriz de información para un conjunto de elección, en el caso del *Status Quo* con no elección está dada por:

$$\begin{bmatrix} X'_i D_i X_i & p_{i,D} \lambda_C V_{i,C} \sum_{j \in \tau_C} p_{ij} x_{ij} \\ p_{i,D} \lambda_C V_{i,C} \sum_{j \in \tau_C} p_{ij} x'_{ij} & p_{i,C} p_{i,D} V_{i,C}^2 \end{bmatrix} \quad (4.129)$$

Considere que el grupo D solo posee una alternativa y ésta tiene asociados atributos todos cero, es decir, se tienen los supuestos de *Status Quo* con no elección. En este caso es claro que  $V_{i,D} = 0$  y por lo tanto la (4.128) se reduce a

$$\begin{bmatrix} 4.121 & 4.123 \\ 4.123' & p_{i,C} p_{i,D} V_{i,C}^2 \end{bmatrix} \quad (4.130)$$

De la ecuación anterior, el elemento dado en (4.123) coincide con su elemento correspondiente en (4.129). De manera similar el elemento  $p_{i,C} p_{i,D} V_{i,C}^2$ . En lo que respecta a (4.121), los elementos correspondientes a las expresiones (4.118)-(4.120) se hacen cero por el supuesto sobre los atributos de la alternativa en D. De esta manera solo quedan los elementos correspondientes a (4.117) y (4.118), los cuales coinciden con los elementos de la expansión del primer elemento de (4.129) ([Goos et al. , 2010](#), consultar apéndice A).

## Parte II

### Optimización de la matriz de información

En la primera parte se desarrolló la matriz de información para el modelo NMNL, lo cual representa el primer objetivo de la tesis. En esta segunda parte se exploran los métodos de optimización para la matriz de información desarrollada desde la perspectiva de la evolución diferencial y un enfoque híbrido entre evolución diferencial y el algoritmo clásico de intercambio de Fedorov.



# 5

## Algoritmos para la estimación de experimentos

### 5.1 Las condiciones para hallar el óptimo

#### 5.1.1 *Condiciones necesarias y suficientes*

El problema planteado en (3.25), aunque en apariencia está bien definido, podría llevar a resultados no esperados, como la no existencia de diseños que satisfagan el criterio de optimalidad deseado o la incapacidad de no poder ser exhibido. Debido a estos problemas es necesario hacer una serie de supuestos adicionales tanto sobre el espacio de búsqueda como de la función objetivo.

#### Condiciones sobre el espacio de búsqueda

Matemáticamente (3.25) podría resultar sin solución debido a la no existencia del mínimo<sup>1</sup> dentro del conjunto de soluciones factibles. Para garantizar la existencia del mínimo es necesario hacer dos suposiciones

- S1 El conjunto de todos los experimentos factibles, denotado por  $\Xi(O)$ , es compacto.
- S2 La matriz de información de Fisher es continua respecto a los puntos del experimento en la región de experimentación ( $O$ ). En otras palabras; en el caso lineal, las funciones base  $f(\mathbf{o})$  son continuas en el caso lineal; mientras que en el caso no lineal  $\frac{\partial \eta(\mathbf{o}, \theta)}{\partial \theta}$  es continua para todo  $\theta$ .

Note que se el supuesto S2 no es más que la petición de que la matriz de información exista. En este sentido, todas las implicaciones de las dos condiciones anteriores se plasman en el siguiente teorema:

**Teorema 5.1** *Si se cumplen las dos condiciones S1 y S2, entonces*

1. *Para cualquier experimento  $\xi$  la matriz de información  $M(\xi)$  es simétrica y definida no negativa.*

---

<sup>1</sup>Si el lector no está familiarizado con la diferencia entre ínfimo y mínimo considere la función  $f(h) = 1/h$  con  $h \in (0, \infty)$ . En este caso  $\min_{(0, \infty)} f(h)$  no existe pero  $\inf_{(0, \infty)} f(h) = 0$ .

2. El conjunto de las matrices de información asociadas a un experimento  $\xi$ , denotado por  $M(O)$  es compacto.
3. Para cualquier matriz  $M \in M(O)$  existe un experimento  $\xi$  con no más de  $m(m+1)/2 + 1$  puntos de soporte con pesos diferentes de cero. Si  $M$  es un punto frontera, entonces el número de puntos soporte es igual a  $m(m+1)/2 + 1$ .

### Condiciones sobre la función objetivo

A diferencia de las condiciones sobre el espacio de búsqueda, las cuales tienen como objetivo garantizar la existencia de una solución, las restricciones sobre la función objetivo buscan que exista una solución práctica del problema planteado por (3.25), así como la implementación de la búsqueda de la solución de una forma más sencilla. De esta manera, el funcional lineal  $\Psi$  debe cumplir

- O1  $\Psi$  es una función convexa, es decir,
- O2  $\Psi$  es una función monótonamente no creciente.
- O3 Sea  $\Xi(q) = \{\xi : \Psi[M(\xi)] \leq q < \infty\}$ . Entonces existe un número real  $q$  tal que el conjunto  $\Xi(q)$  no es vacío.
- O4 Para cualquier par de diseños  $\xi$  y  $\bar{\xi}$  tales que  $\xi \in \Xi(q)$  y  $\bar{\xi} \in \Xi$  se cumple<sup>2</sup>

$$\Psi[(1-\alpha)M(\xi) + \alpha M(\bar{\xi})] = \Psi[M(\xi)] + \alpha \int_X \psi(\mathbf{o}, \xi) \bar{\xi}(d\mathbf{o}) + h(\alpha|\xi, \bar{\xi}) \quad (5.1)$$

donde  $\lim_{\alpha \rightarrow 0} \frac{h(\alpha|\xi, \bar{\xi})}{\alpha} = 0$  y  $\psi(\mathbf{o}, \xi)$  es una función que satisface la ecuación anterior.

Las condiciones O1 y O2 son condiciones que satisfacen todos los criterios de optimidad (Fedorov, 2013), mientras que la condición O3 pide la existencia de valores no infinitos para que la comparación y la optimización tengan sentidos computacionalmente y la condición O4 da pauta para la búsqueda del mínimo.

#### 5.1.2 Un objetivo equivalente

Las condiciones anteriores tienen como consecuencia siguiente el teorema:

**Teorema 5.2** *Si las condiciones S1, S2 y O1-O4 se cumplen entonces*

1. Existe un experimento óptimo  $\xi^*$  tal que no contiene más de  $m(m+1)/2$  puntos de soporte.
2. El conjunto de experimentos óptimo es convexo.

---

<sup>2</sup>La condición se refiere a la existencia de la derivada direccional, la cual recibe el nombre de derivada de Gâteaux debido a que ésta en un espacio topológico diferente del euclídeo. Esta definición de derivada conserva las propiedades de la regla de la cadena y del teorema de Taylor.

3. Una condición necesaria y suficiente para que un experimento  $\xi^*$  sea óptimo es que satisfaga

$$\min_{\mathbf{o} \in O} \psi(\mathbf{x}, \xi) \geq 0 \quad (5.2)$$

4. La función  $\psi$  es igual a cero en casi todo punto<sup>3</sup> en el sup( $\xi^*$ )

Note que la propiedad 4 del Teorema 5.2 es consecuencia de la compacidad y de que el mÁnimo exista. Si  $\psi$  no fuera cero en casi todo punto, significaría que existe un conjunto no finito de puntos  $\mathbf{x}$  tales que son diferentes de cero y entonces  $\min_{\mathbf{x} \in X} \psi(\mathbf{x}, \xi) \neq \inf_{\mathbf{x} \in X} \psi(\mathbf{x}, \xi)$ . Por otro lado el mismo teorema muestra que el problema planteado en (3.25) es equivalente al problema planteado en (5.2). Mientras que la condición O4 exhibe una forma sencilla de calcular  $\psi$ . De esta manera se tiene

$$\lim_{\alpha \rightarrow 0} \frac{\Psi[(1 - \alpha)M(\xi) + \alpha M(\bar{\xi})] - \Psi[M(\xi)]}{\alpha} = \int_X \psi(\mathbf{x}, \xi) \bar{\xi}(d\mathbf{x}) \quad (5.3)$$

**Un ejemplo del uso de la expresión (5.3)**

La expresión (5.3) permite desarrollar funciones equivalentes a los criterios de optimidad. Por ejemplo, definiendo

$$M_\alpha(\xi, \bar{\xi}) = (1 - \alpha)M(\xi) + \alpha M(\bar{\xi})$$

y denotando a la función definida en Elemento 5.1 por  $G'(\alpha)$ , pues es la derivada de Gâteaux. Entonces para el criterio D-óptimo:

$$\begin{aligned} G'(\alpha) &= \frac{\partial \log |M_\alpha|}{\partial \alpha} = \text{traza} \left[ M_\alpha^{-1} \frac{dM_\alpha}{d\alpha} \right] \\ &= \text{traza} [M_\alpha^{-1} (M(\bar{\xi}) - M(\xi))] \\ &\quad \text{Al aplicar el límite como en (5.3) se tiene} \\ \lim_{\alpha \rightarrow 0} G'(\alpha) &= \text{traza} [M^{-1}(\xi) (M(\bar{\xi}) - M(\xi))] \\ &= p - \text{traza} [M^{-1}(\xi) M(\bar{\xi})] \\ &\quad \text{Luego, usando (3.14) se tiene que} \\ &= p - \text{traza} \left[ M^{-1}(\xi) \int_O \mu(\mathbf{o}) \bar{\xi}(d\mathbf{o}) \right] \\ &= \int_O [p - \text{traza} [M^{-1}(\xi) \mu(\mathbf{o})]] \bar{\xi}(d\mathbf{o}) \end{aligned}$$

Recuerde que (3.14) es una forma explícita de la integral de Lebesgue, y así se obtiene un criterio equivalente para el criterio D-óptimo dado por

$$\psi(\mathbf{x}, \xi) = p - \text{traza} [M^{-1}(\xi) \mu(\mathbf{o})]$$

---

<sup>3</sup>Esto significa que sólo existe un conjunto finito donde la función no es cero.

## 5.2 Un algoritmo de primer orden

El algoritmo empleado para la obtención de diseños óptimos por tradición es el “*Algoritmo de intercambio de Fedorov*” ([Atkinson et al. , 2007](#)), el cual es la base de los algoritmos que se encuentran en software como lo son SAS para la estimación de experimentos óptimos ([SAS, 2013](#)). El algoritmo se basa en dos hechos fundamentales

1. Dado un diseño de  $N$  puntos  $\xi_N$ , el punto  $N + 1$  está dado por

$$\mathbf{o}_{N+1} = \arg \min_{\mathbf{o}} \psi(\mathbf{o}, \xi_N) = \arg \max_{\mathbf{o}} \varphi(\mathbf{o}, \xi_N) \quad (5.4)$$

donde  $\varphi = -\psi + C$  y es llamada función de sensibilidad.

2. Dado  $\mathbf{o}_{N+1}$  y  $\xi_N$  entonces<sup>4</sup>

$$\xi_{N+1} = (1 - \alpha) \xi_N + \alpha_N \xi(\mathbf{o}_{N+1}) \quad (5.5)$$

donde  $\alpha$  puede ser seleccionado usando varios criterios, por ejemplo

$$\alpha_N = \begin{cases} \alpha_{N-1} & \text{si } \Psi[(1 - \alpha_{N-1})\xi_N + \alpha_{N-1}\xi(\mathbf{o}_{N+1})] < \Psi[\xi_N] \\ \gamma\alpha_{N-1} & \text{en otro caso con } 0 < \gamma < 1 \end{cases} \quad (5.6)$$

### 5.2.1 Pasos hacia atrás y hacia adelante

Aunque (5.4) garantiza encontrar un diseño, debido a que la actualización de los pesos hará que los puntos no útiles tengan pesos nulos, los pasos hacia adelante y hacia atrás permitirán mejorar la velocidad de convergencia o en el caso particular de los modelos para la toma de decisión, donde los pesos no se ven involucrados en la matriz de información (ver Capítulo 4), permite determinar qué conjuntos de elección son los mejores para alcanzar el objetivo.

La idea del paso hacia adelante consiste en la implementación de (5.4); mientras que la idea del paso hacia atrás consiste en la eliminación del peor punto en el diseño con base en la función de sensibilidad. De esta manera, el punto candidato a retirar del diseño está dado por

$$\mathbf{o}_{N+1}^- = \arg \min_{\mathbf{o} \in X_N} \varphi(\mathbf{o}, \xi_N) \quad (5.7)$$

y por consiguiente, si  $\mathbf{o}_{N+1}^+ = \arg \min_{\mathbf{o}} \varphi(\mathbf{o}, \xi_N)$  entonces

$$\mathbf{o}_{N+1} = \arg \max [\mathbf{o}_{N+1}^-, \mathbf{o}_{N+1}^+] \quad (5.8)$$

En otras palabras, el punto que más aporta debe ser agregado al diseño.

---

<sup>4</sup> $\xi_{N+1}$  es resultado de una combinación convexa entre el experimento del paso  $N$  y el experimento con el punto determinado por la (5.4).

## 5.3 La evolución diferencial: una alternativa

### 5.3.1 ¿Qué es la evolución diferencial?

Considere la circunstancia en donde un grupo de personas (población), que individualmente exploran nuevas áreas, profundizan sobre un tema para después comparar y validar el conocimiento encontrado con el resto de la población, permitiendo a la población acercarse cada vez más a la solución de un problema específico, está circunstancia es la inspiración de la evolución diferencial (DE). Feoktistov (2006) define a la DE como un modelo matemático sencillo y simple que trata de imitar un proceso complejo al que se conoce como evolución<sup>5</sup> y el cual está dado por

$$u_{p,d}(t+1) = s_{r_1,d}(t) + F(s_{r_2,d} - s_{r_3,d}) \quad (5.9)$$

donde  $u_{p,d}(t+1)$  es el elemento  $d$ -ésimo del individuo  $p$  en la población, resultado de una evolución vista como una combinación lineal entre tres elementos aleatorios  $s_{r_i,d}$  de la generación actual ( $t$ ) de la población actual, el cual es candidato a mejorar la población (soluciones) en la siguiente generación ( $t+1$ ). Dicha combinación involucra una tasa de evolución  $F$  y  $t$  representa la generación de la población.

La forma en que se definió la evolución en (5.9) obliga a que exista un conjunto de soluciones iniciales (población). En la práctica, esas soluciones se eligen de manera aleatoria dentro de una cierta región. En general, la idea de evolución permite resolver una gran cantidad de problemas, en especial problemas complejos que impliquen combinatoria o modelos no lineales (Fig. 5.1). Matemáticamente la evolución diferencial permite resolver problemas

$$\min f(\mathbf{s}) \quad (5.10)$$

Particularmente, Paredes-García y Castaño Tostado (2015) mostraron que es posible ajustar la idea de la DE para resolver (3.25), es decir, la obtención de diseños experimentales óptimos.

### 5.3.2 ¿Cómo funciona la evolución diferencial?

Considere el problema

$$\min f(\mathbf{s}) = \min (s_1^2 + s_2^2) \quad \text{con } \mathbf{s} = [s_1, s_2]' . \quad (5.11)$$

un problema donde solución del mínimo global es  $\mathbf{s} = (0, 0)$ . Usando la idea de un conjunto inicial de soluciones que evoluciona de acuerdo a (5.9) se puede aproximar la solución al problema antes mencionado (Fig. 5.2).

De manera genérica, los algoritmos evolutivos diferenciales, constan de 4 etapas fundamentales:

1. **Inicialización.** La inicialización consiste en la determinación de las soluciones (población) iniciales  $\mathbf{x}$  de dimensión  $D$  y de tamaño<sup>6</sup>  $NP \geq 4$  en una región de búsqueda determinada, aunque ésta no necesariamente debe contener a la solución (Fig. 5.3). La manera más práctica de determinar esta población consiste en tomar  $NP$  de manera aleatoria dentro de la región de búsqueda inicial.

<sup>5</sup>Algunos autores le llaman mutación por la relación entre la evolución diferencial y los algoritmos genéticos.

<sup>6</sup>La desigualdad es consecuencia de (5.9).

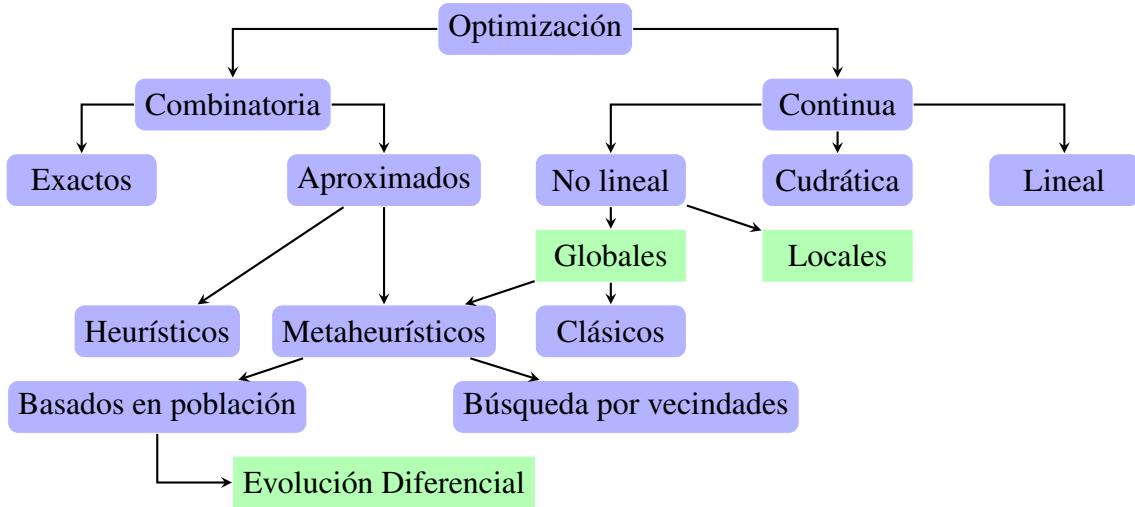


Figura 5.1: Tipos de problemas de optimización y posibles métodos de soluciones (Feoktis-tov, 2006). En

2. Mutación. La mutación, es la evolución del conjunto de soluciones empleando (5.9), la cual representa la manera clásica de mutación entre los individuos de una población (Price *et al.*, 2005), aunque existen variaciones de ésta como la empleada por Li y Zhang (2014).
3. Cruzamiento. Es una etapa complemento de la mutación y que únicamente aplica para problemas de dimensión mayor a 1. Consiste en una acción para determinar de manera aleatoria si la mutación ocurre o no en cada uno de los elementos que conforman a cada una de las soluciones del conjunto. De esta manera cada uno de los elementos finales  $v_{p,d}$  de cada solución, candidatos a mejorar al conjunto solución en la siguiente generación está dado por

$$v_{p,d}(t+1) = \begin{cases} u_{p,d}(t+1) & \text{si } U \sim U(0,1) < CR \\ x_{p,d}(t) & \text{en otro caso.} \end{cases} \quad (5.12)$$

donde CR es la tasa de cruzamiento. Cuando esta etapa es omitida, se considera que siempre existe cruzamiento y es equivalente a elegir CR = 1.

4. Selección. La selección es la comparación entre  $v(t+1)$  y  $x(t)$  y la elección de la siguiente generación del conjunto de soluciones mediante la siguiente ecuación

$$s_p(t+1) = \begin{cases} v_p(t+1) & \text{si } f(v_p) < f(s_p) \\ s_p(t) & \text{en otro caso.} \end{cases} \quad (5.13)$$

### 5.3.3 Las ventajas y desventajas de la evolución diferencial

El Teorema 5.2 muestra la necesidad de la compacidad y la convexidad; la compacidad para asegurar que el mínimo está dentro de la región de búsqueda y la convexidad para poder asegurar que cada vez que se emplee (5.5),  $\xi_{N+1}$  seguirá estando en la región de

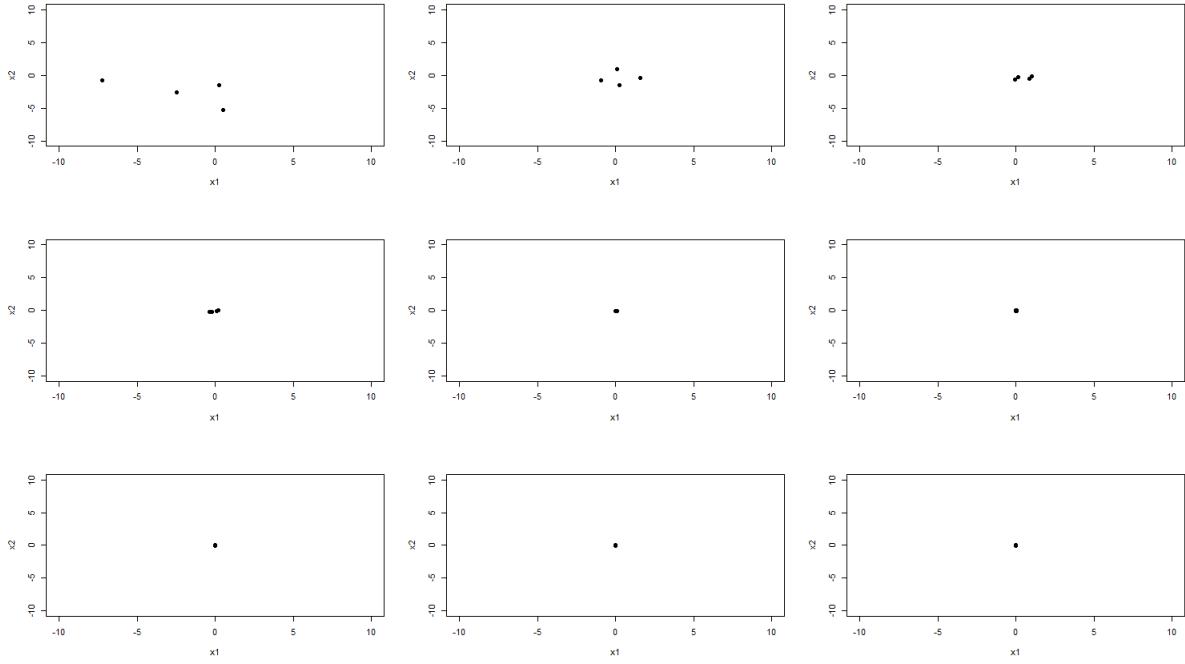


Figura 5.2: Evolución de la población de soluciones usando como región de búsqueda el espacio  $[-10, 10] \times [-10, 10]$ . En la parte superior izquierda la solución inicial. De izquierda a derecha y de arriba a abajo la solución después de 5, 10, 15, 20, 25, 30, 35 y 40 iteraciones.

búsqueda. Y por lo tanto, la combinación de ambas condiciones garantiza la efectividad del “Algoritmo de intercambio de Fedorov”.

Al emplear DE, por el contrario, no es necesario ni la compacidad ni la convexidad para encontrar un experimento. La falta de compacidad, que elimina seguridad de la existencia del mínimo, es sobrelevada por la misma filosofía de la metaheurística donde el objetivo no es encontrar la mejor solución sino una buena solución; de esta manera, si se considera la función

$$\min_{x \in (0, \infty)} f(x)$$

donde se sabe que el mínimo no existe, el empleo de DE nos llevará a soluciones cercanas a cero que resultan útiles en la práctica. Por otro lado, el empleo la búsqueda en regiones no convexas puede ser sorteado empleado funciones de ajuste penalizadas ([Schneider y Krohling, 2014](#); [Li y Zhang, 2014](#)) con el fin de sancionar el ajuste de las soluciones, y en nuestro caso diseños, que están fuera de la región deseada.

Aunque la falta de compacidad y convexidad pueden ser sorteadas por la DE cuando no están presentes éstas, una restricción importante es la determinación de diseños con un mínimo de puntos. El Teorema 5.2, aparte de asegurar la existencia de una solución como el uso del algoritmo de Fedorov, proporcionaba, en su primer punto, una condición de paro para la determinación del diseño óptimo con un mínimo de puntos. Al no tener las condiciones con el empleo de la DE, lo mejor que se podrá determinar será un buen experimento de N puntos bajo un criterio específico y en caso de no existir todos los experimentos tendrán un

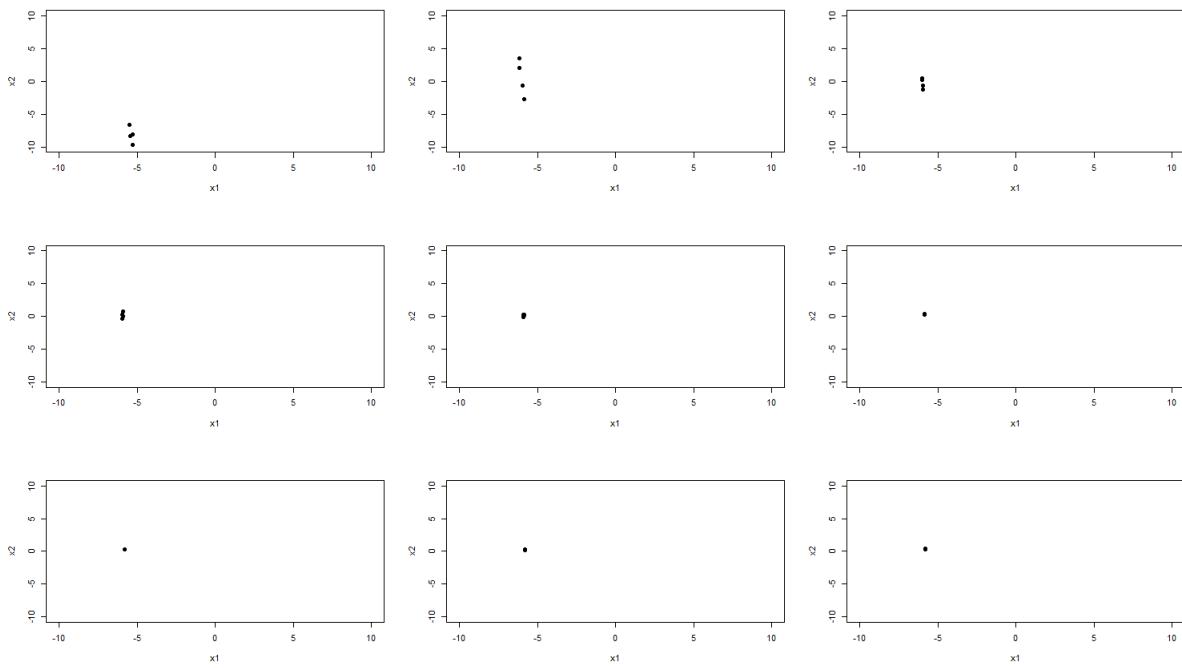


Figura 5.3: Evolución de la población de soluciones usando como región de búsqueda el espacio  $[-10, -5] \times [-10, -5]$ . En la parte superior izquierda la solución inicial. De izquierda a derecha y de arriba a abajo la solución después de 5, 10, 15, 20, 25, 30, 35 y 40 iteraciones.

mismo valor de optimalidad (veáse Capítulo 7). Sin embargo, la ausencia de convexidad y compacidad pueden llevar a diseños factibles en regiones en regiones sin esas propiedades recordando que éstas son necesarias para poder tener una manera de construir diseños y no, que los diseños sean consecuencia de dichas propiedades.

# 6

## Implementación del algoritmo basado de DE

El propósito de este capítulo es mostrar los detalles para la implementación el algoritmo propuesto por [Paredes-García y Castaño Tostado \(2015\)](#).

### 6.1 La función objetivo

El primer paso para la implementación del algoritmo es la definición de la función objetivo. Como se mencionó en el capítulo anterior, a diferencia del “*Algoritmo de intercambio de Fedorov*”, para emplear la DE no es necesario definir una función convexa. Debido a esto, para este primer paso es posible seguir dos caminos; usar la definición propia del criterio o usar la función de sensibilidad del criterio mismo. Por lo tanto para el criterio D-óptimo es posible emplear cualquiera de las siguientes funciones objetivo

$$\arg \min_{\xi} |M^{-1}| \quad (6.1)$$

$$-\arg \max_{\xi} |M| \quad (6.2)$$

$$\arg \min_{\xi} \max_{\mathbf{x}} \text{traza} [\mu(\mathbf{x}) M] \quad (6.3)$$

Mientras que para el criterio A-óptimo si matriz A es la identidad entonces se tiene que se puede emplear

$$\arg \min_{\xi} \text{traza} [M^{-1}] \quad (6.4)$$

$$\arg \min_{\xi} \max_{\mathbf{o}} \text{traza} [\mu(\mathbf{x}) M^2] \quad (6.5)$$

Al usar como función objetivo cualquiera de las expresiones anteriores es posible construir un diseño óptimo bajo el criterio D () o el criterio A (). Sin embargo, como se hizo notar en el Capítulo 3 la matriz de un modelo no lineal, como lo es el modelo NMNL, depende de parámetros desconocidos, en particular de  $\beta$  y  $\lambda$  para el ejemplo mencionado. [Fedorov \(2013\)](#) menciona dos enfoques principales: diseños locales o diseños bayesianos. El primero consiste en dar una configuración fija inicial para los parámetros desconocidos en la matriz de información; el segundo, por otro lado consiste en suponer que existe una distribución de probabilidad sobre los parámetros. A lo largo de la tesis se trabajará con diseños locales.

## 6.2 Especificación del algoritmo

Como ya se mencionó con el solo empleo de las 4 etapas básicas de la DE para la optimización de la (6.1) - (6.5) es posible dar una buena estimación de un experimento óptimo, como el mostrado en la Tabla 6.1. Aunque pueden emplearse métodos para mejorar la velocidad de convergencia del algoritmo, como el propuesto por [Paredes-García y Castaño Tostado \(2015\)](#) que emplea búsqueda local usando combinaciones convexas para hacer un paso hacia atrás o el propuesto por [Li y Zhang \(2014\)](#) que basado en usar un diseño factorial para determinar los efectos de los elementos de cada individuo.

Tabla 6.1: Ejemplo de experimento D-óptimo local para 3 conjunto de elección con 3 alternativas de 3 niveles cada uno. La jerarquía determinada por el atributo 1 (diferente de 3 e igual a 3).

Conjunto de elección	Experimento D-óptimo local		
	$\beta = [1, 1, 1]$		
	$\lambda_C = 0.8$		
	$\lambda_D = 0.1$		
1	1	3	2
	2	2	1
	3	2	1
2	1	2	1
	2	3	1
	3	1	3
3	1	2	1
	2	2	2
	3	3	1

### 6.2.1 La codificación de la información

En el algoritmo se propone la codificación empleando un arreglo de matrices, donde cada matriz posee una dimensión de  $2 \times D$  y representa a un diseño. De manera que se pueda trabajar con una población de diseños. Además de considerar a cada diseño un elemento de dimensión uno con el fin de eliminar el cruzamiento. Sin embargo, la obtención de diseños para el modelo MNL o NMNL depende de la determinación de un conjunto de conjuntos de elección, donde en cada uno de los conjuntos de elección aparece un cierto número de alternativas delineadas por una configuración específica de un conjunto de atributos. Por lo tanto, para poder codificar la información se pueden emplear dos soluciones sumamente fáciles de implementar

1. En software como Matlab es posible definir hipermatrices. Entonces para manejar a la población es posible acomodarla en un arreglo de matrices, donde cada matriz contiene a todos los conjuntos de elección y después un método para partir la matriz en los conjuntos de elección para la evaluación experimental de la matriz de información.

2. En software como R que no puede manejar hipermatrices; pero que puede manejar listas, entonces la implementación puede emplear listas de listas. Donde la primera lista contiene una lista de N matrices que representan a la matriz de información.

### 6.2.2 Los parámetros del algoritmo

Una parte importante son los parámetros de ajuste del algoritmo; el tamaño de la población, el factor de mutación y la tasa de cruzamiento. En lo que respecta al tamaño de población los autores dejan en claro que si bien una población de mayor tamaño requiere un número menor de iteraciones, incluso empleando el mínimo de población ( $NP = 4$ ) la estimación de los experimentos se hace en muchos casos en menos de 50 iteraciones. Mientras en lo que concierne al factor de mutación especifican usar valores cercanos a 1 debido a que una tasa de mutación pequeña puede hacer que sean requeridas un número grande de iteraciones o provocar que el algoritmo se quede en una solución no tan buena (Fig. 6.1). Finalmente, para la tasa de cruzamiento, debido a que se considera en la representación un problema de dimensión uno (como se mencionó al principio de esta sección), entonces la tasa de cruzamiento es 1.

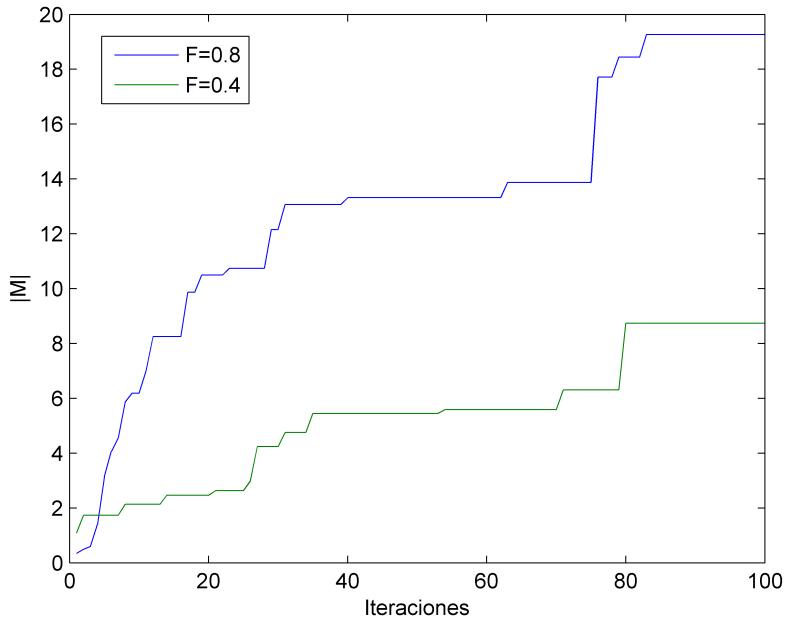


Figura 6.1: Comparación de la evolución de experimentos D-óptimos para el modelo NMNL para diferentes valores de  $F$ .

Otra parte a especificar son las características del diseño; el número de alternativas y el número de conjuntos elección. Para el número de conjuntos de elección ( $n$ ) está dado por

$$n = \arg \min_N \Psi [M(\xi_N)] > 0 \quad (6.6)$$

ya que mientras el número de puntos en el experimento sea insuficiente el criterio de optimidad será cero. Mientras, en el caso del número de alternativas que conforman a cada

conjunto de elección, depende del modelo. En el caso del modelo MNL dos alternativas por conjunto de elección son suficientes para establecer un orden de preferencias. Sin embargo, para el caso del modelo NMNL, si bien dos alternativas (una de cada nido) por conjunto de elección permiten que la probabilidad de elegir algún nido ((4.21) y (4.22)) no sea nula; este número de alternativas no evita que la regla de decisión sea más compleja que lo especificado por modelo MNL (Fig. 6.2), matemáticamente, por vacuidad, no existen diferencias entre las alternativas de cada nido y por lo tanto

$$\lambda_C = \lambda_D = 0$$

y así de acuerdo a lo mencionado Sección 2.2 se vuelve al caso del modelo MNL.

La última parte consiste en especificar la configuración inicial del modelo, debido a que la matriz de información depende de parámetros desconocidos (Sección 2.2) y, una solución a ello es la obtención de diseños locales dada una configuración inicial. En el caso del modelo MNL no hay ninguna restricción sobre los valores a elegir. Sin embargo, en el caso del modelo NMNL no todas las configuraciones son válidas, el lector debe recordar que si un nido posee una única alternativa entonces el parámetro  $\lambda$  asociado debe tener un valor nulo; de lo contrario, la matriz de información dejaría de ser definida positiva.



Figura 6.2: A la izquierda la regla de decisión impuesta por el modelo MNL y a la derecha la regla de decisión impuesta por el modelo NMNL cuando solo posee una alternativa por nido.

## 6.3 El algoritmo del intercambio de Fedorov usando DE

A esta altura se poseen todas las herramientas necesarias para la obtención de diseños usando evolución diferencial. Sin embargo, no se ha mencionado hasta el momento un método para optimizar la (5.4). La realidad es que depende del modelo. A pesar de esto, una forma es emplear evolución diferencial para la optimización de la parte crucial del algoritmo de Fedorov. De esta manera si el lector desea obtener un diseño con  $N$  conjuntos de elección debe seguir los siguientes pasos

1. Eligir  $N$  conjuntos de elección aleatorios dentro de la región de búsqueda.
2. Empleando DE optimizar la (5.4) para estimar un nuevo conjunto de elección y agréguelo al diseño.
3. De entre todos los conjuntos de elección determinar el conjunto de elección que menos aporte y sacarlo del diseño.
4. Repetir los pasos hasta que la mejora no sea significativa de acuerdo a un parámetro de tolerancia.

## 6.4 Construcción de diseños y rendimiento del algoritmo

Para evaluar al algoritmo, se contrastó el diseño encontrado por el algoritmo DE y el hallado por [Goos et al. \(2010\)](#) (Tabla 6.2). Al observar la alternativas creadas para cada uno de los experimentos se puede observar que el diseño de Goos et al. posee alternativas contrastantes debido a que los cambios son en una gran cantidad de atributos; mientras que las alternativas del diseño, obtenido con el algoritmo propuesto, poseen cambios menos visibles debido a que estos son en uno o dos atributos.

Tabla 6.2: Diseño D-óptimo local para el modelo NMNL con No-Choice para 8 conjuntos de elección con 6 atributos de 2 niveles, considerando que los atributos poseen un efecto negativo y con un parámetro de disimilitud entre las alternativas reales de 0.5. A la izquierda se observa el diseño obtenido por [Goos et al. \(2010\)](#) y a la derecha el obtenido usando el algoritmo propuesto por [Paredes-García y Castaño Tostado \(2015\)](#). Note que  $\lambda_D$  no aparece debido a que solo posee una alternativa y por tanto su valor es fijo.

Conjunto de elección	Experimento D-óptimo local $\beta = -[1, 1, 1, 1, 1, 1]$ $\lambda_C = 0.5$						Conjunto de elección	Experimento D-óptimo local $\beta = -[1, 1, 1, 1, 1, 1]$ $\lambda_C = 0.5$					
	1	2	2	1	2	2		1	1	2	1	1	1
<b>1</b>	2	2	1	2	1	2	<b>1</b>	1	1	1	2	1	2
	0	0	0	0	0	0		0	0	0	0	0	0
	0	0	0	0	0	0		0	0	0	0	0	0
<b>2</b>	2	1	1	2	2	2	<b>2</b>	1	1	1	1	1	2
	2	2	2	2	1	1		2	2	1	1	1	1
	0	0	0	0	0	0		0	0	0	0	0	0
<b>3</b>	2	2	1	1	1	2	<b>3</b>	1	1	2	1	1	1
	1	1	2	2	2	1		1	2	1	2	1	1
	0	0	0	0	0	0		0	0	0	0	0	0
<b>4</b>	2	2	1	1	2	1	<b>4</b>	1	1	1	1	2	1
	1	1	2	2	1	2		1	1	1	2	1	1
	0	0	0	0	0	0		0	0	0	0	0	0
<b>5</b>	1	2	1	2	2	1	<b>5</b>	1	2	1	1	1	2
	2	1	2	1	1	2		1	1	1	2	2	1
	0	0	0	0	0	0		0	0	0	0	0	0
<b>6</b>	1	2	1	2	1	2	<b>6</b>	1	2	1	1	1	1
	2	1	1	1	2	1		1	1	2	1	2	2
	0	0	0	0	0	0		0	0	0	0	0	0
<b>7</b>	2	1	1	2	2	1	<b>7</b>	1	2	2	1	1	1
	1	2	2	1	2	2		2	1	1	1	1	1
	0	0	0	0	0	0		0	0	0	0	0	0
<b>8</b>	2	1	2	2	2	1	<b>8</b>	1	1	1	1	2	1
	1	2	2	1	1	2		2	1	2	1	1	1
	0	0	0	0	0	0		0	0	0	0	0	0

En la Tabla 6.3 se pueden observar las probabilidades de las alternativas de ambos diseños para cada uno de sus conjunto de elección. En ella se pueden observar dos estrategias distintas. El diseño propuesto por Goos *et al.* posee alternativas, dentro del conjunto de elecciones reales, con probabilidades iguales; en otras palabras, el experimento busca responder la pregunta: “*¿Cuáles atributos permiten el cambio en un empate hipotético?*” Por otro lado, el experimento estimado por el algoritmo propuesto posee alternativas, dentro del conjunto de elecciones reales, principalmente extremas; una *altamente probable* y otra *poco probable*, una vez que se ha elegido seleccionar una alternativa real, es decir, busca responder la pregunta: “*¿Qué cambio permitió a algo poco probable ser elegido?*”. Ambas estrategias en apariencia deberían permitir la estimación de atributos significativos en la toma de decisión. Por un lado los atributos que inclinan la balanza a favor de una alternativa deberían ser significativos y por otro lado los atributos que permitieron a una alternativa no muy atractiva ser elegida deberían ser importantes.

Tabla 6.3: Tabla de comparación de los diseños expuestos en la Tabla 6.2 vía modelo.

CE	Experimento Alg. P.			Experimento Goos et al.		
	Pr(A)	Pr(B)	Pr(SQ)	Pr(A)	Pr(B)	Pr(SQ)
<b>1</b>	0.0249	0.0092	0.9659	0.0047	0.0047	0.9906
<b>2</b>	0.0249	0.0092	0.9659	0.0047	0.0047	0.9906
<b>3</b>	0.0249	0.0092	0.9659	0.0077	0.0077	0.9846
<b>4</b>	0.0205	0.0205	0.959	0.0077	0.0077	0.9846
<b>5</b>	0.0126	0.0126	0.9748	0.0077	0.0077	0.9846
<b>6</b>	0.0275	0.0037	0.9688	0.0077	0.0077	0.9846
<b>7</b>	0.0092	0.0249	0.9659	0.0094	0.0034	0.9872
<b>8</b>	0.0249	0.0092	0.9659	0.0094	0.0034	0.9872
$ M^{-1} $	$2.5269 \times 10^8$			$1.6125 \times 10^{10}$		

\*CE Conjunto de elección

\*SQ *Status Quo*

Pero, ¿cuál estrategia es mejor? Para la comparación de las estrategias, una manera absoluta es al comparar  $|M^{-1}|$ , el cual puede observarse en la misma Tabla 6.2, donde el volumen del elipsoide creado por la matriz de información del diseño obtenido con el algoritmo propuesto es menor al diseño obtenido por Goos *et al.* Otra forma de comparar es al observar las probabilidades asociadas al *Status Quo*, donde el diseño obtenido con el algoritmo propuesto posee probabilidades más pequeñas que el diseño obtenido por Goos *et al.* Debido a que la elección del *Status Quo* significa pérdida de información, entonces, el experimento estimado con el algoritmo propuesto tiene más oportunidad de recabar información. Sin embargo, uno de los problemas expuestos por Louviere *et al.* (2008) es la consistencia en la respuesta y ésta puede verse afectada si el usuario no puede percibir diferencias con tanta facilidad. En este aspecto, el experimento de Goos *et al.* puede tener una ventaja debido a diferencias más marcadas en las alternativas; aunque es necesario hacer notar que vía modelo, las alternativas creadas por el algoritmo propuesto deberían hacer más fácil la disyuntiva del decisor.

# 7

## Diseños óptimos locales para el modelo NMNL

En el Capítulo 6 se comparó el rendimiento del algoritmo propuesto con base en un diseño ya antes generado. Sin embargo, aún no se ha hablado de los efectos de los parámetros iniciales ( $\beta, \lambda_C, \lambda_D$ ), más allá de su significado teórico. En este sentido, como conclusión de esta segunda parte este capítulo muestra cómo el significado de los parámetros se ve reflejado en los diseños creados.

Para plasmar los efectos de estos parámetros, se consideró un experimento para el modelo NMNL. El conjunto de diseños consistió de 4 alternativas de 6 atributos, cada atributo con 4 niveles y donde cada conjunto perteneciente al nivel superior de la jerarquía posee 2 alternativas; además se propuso que la jerarquía sólo dependiera del primer atributo, donde los parámetros  $\lambda_C, \lambda_D$  y  $\beta$  se movieron según la Tabla 7.1

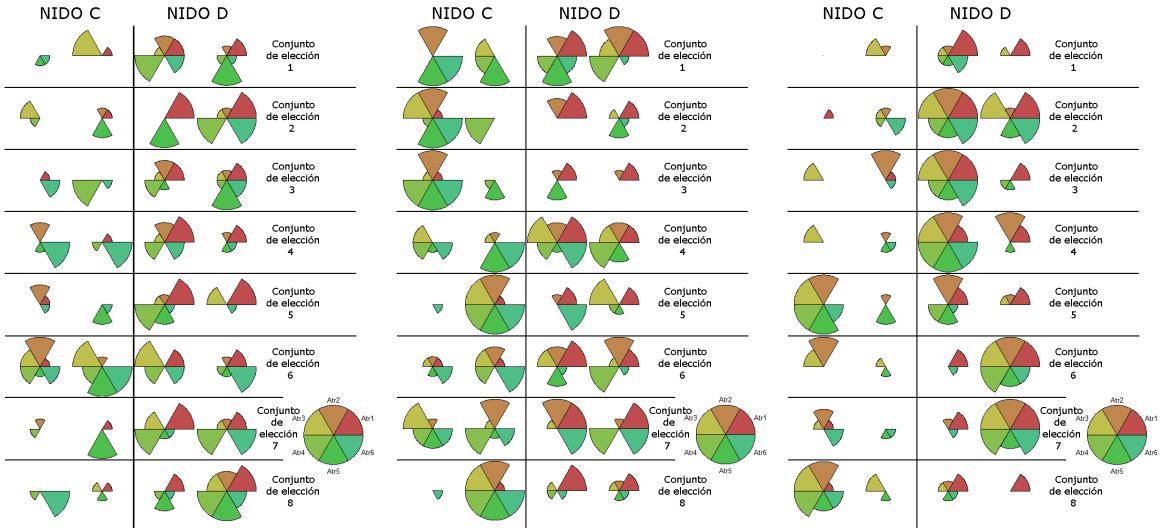
Tabla 7.1: Tabla de configuración de parámetros para mostrar los efectos de los mismos en los diseños.

Diseño.	a	b	c	d	e	f	g	h
$\lambda_C$	0.2	0.2	0.8	0.8	0.2	0.2	0.8	0.8
$\lambda_C$	0.1	0.9	0.1	0.9	0.1	0.9	0.1	0.9
$\beta$	$\mathbf{1}_6$	$\mathbf{1}_6$	$\mathbf{1}_6$	$\mathbf{1}_6$	$-\mathbf{1}_6$	$-\mathbf{1}_6$	$-\mathbf{1}_6$	$-\mathbf{1}_6$

### 7.1 Comparación gráfica

Aunque muchos efectos de los parámetros y las restricciones son vistas más claramente vía modelo, los parámetros de disimilitud,  $\lambda_C, \lambda_D$ , sí pueden ser apreciados gráficamente. Al ser parámetros de disimilitud entre las alternativas del mismo nido, se esperaría que las alternativas plasmadas en los experimentos fueran diferentes entre sí, cuanto más cercanos los  $\lambda_i$ 's estuvieran de uno; y más iguales entre sí, cuanto más cercanos  $\lambda_i$ 's estuviera de cero. En la Fig. 7.1 se puede observar primeramente que no aparecen los diseños correspondientes a la configuración  $\lambda_C = 0.2$  y  $\lambda_D = 0.1$ . En la Fig. 7.1 se puede observar que las diferencias de las alternativas de cada conjunto de elección perteneciente a cada diseño está influenciada por los  $\lambda_i$ 's. Describiendo diseño por diseño, como mero ejercicio confirmatorio:

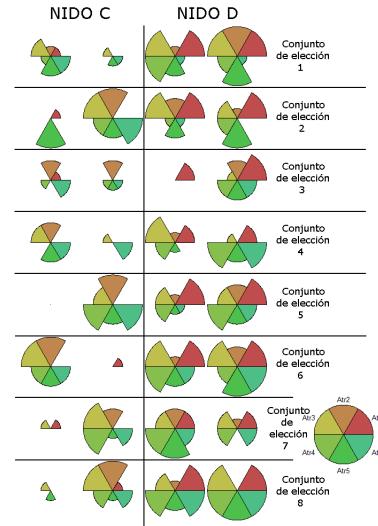
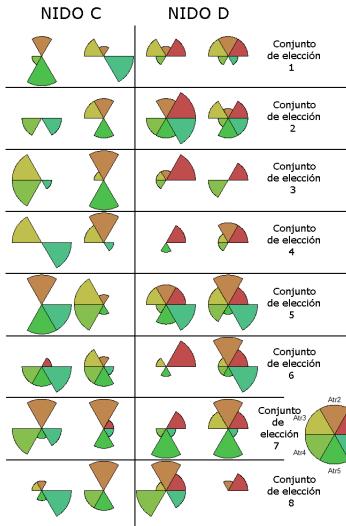
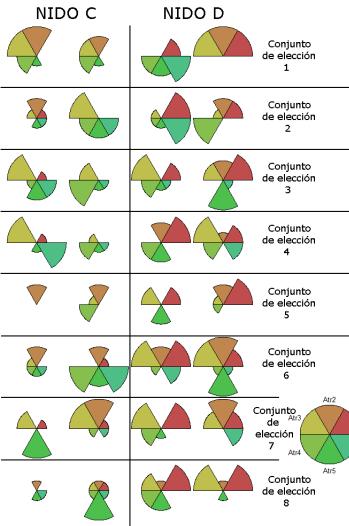
b. Se puede observar en la figura correspondiente al diseño b de la Tabla 7.1, que las similitudes entre las alternativas de los conjuntos de elección del nido D sólo es entre 0 y 2 atributos; mientras que las similitudes entre las alternativas de los conjuntos de elección del nido C es de 1 hasta 3 atributos. Lo cual corresponde a lo esperado.



(b)  $\beta = \mathbf{1}_6, \lambda_C = 0.2$  y  $\lambda_D = 0.9$

(c)  $\beta = \mathbf{1}_6, \lambda_C = 0.8$  y  $\lambda_D = 0.1$

(d)  $\beta = \mathbf{1}_6, \lambda_C = 0.8$  y  $\lambda_D = 0.9$



(f)  $\beta = -\mathbf{1}_6, \lambda_C = 0.2$  y  $\lambda_D = 0.9$  (g)  $\beta = -\mathbf{1}_6, \lambda_C = 0.8$  y  $\lambda_D = 0.1$  (h)  $\beta = -\mathbf{1}_6, \lambda_C = 0.8$  y  $\lambda_D = 0.9$

Figura 7.1: Gráficos de estrella para los 6 experimentos creados, donde para cada imagen, las columnas representan las alternativas 1,2,3 y 4 y las filas corresponden de arriba a abajo a cada uno de los conjuntos de elección. Además las primeras dos alternativas corresponden al conjunto C y las otras dos al conjunto D. Los atributos que no aparecen coloreados se considera que están en el nivel 1.

c. Similarmente en la figura correspondiente al diseño c de la Tabla 7.1, se puede ver que las similitudes entre las alternativas de los conjuntos de elección del nido D van de uno

a cinco atributos<sup>1</sup>; mientras que las similitudes entre las alternativas de los conjuntos de elección del nido C es de nula a una hasta una correspondencia de 2 atributos. Lo cual sigue con la línea marcada por la configuración de los parámetros de disimilitud.

- d. Igualmente en la figura correspondiente al experimento d de la Tabla 7.1, se puede ver un comportamiento parecido en la cantidad de similitudes entre los atributos de los conjuntos de elección de ambos nidos. Ambos poseen la mayoría de sus similitudes debajo de 1 y un sólo un conjunto de elección en ambos casos con una similitud de 4; aunque esto no es extraño, ya que el parámetro es un límite de hasta cuán tan diferentes son las elecciones y no significa que siempre deban serlo así si no es conveniente para el objetivo del criterio.
- f - h. En estos diseños se ve un comportamiento respecto a su configuración recíproca de los diseños b, c y d de la Tabla 7.1. Por lo que el efecto de la positividad o negatividad de los atributos que refleja  $\beta$  no es visible.

Por otro lado, repasando la Tabla 7.1 se aprecia que hay ocho configuraciones a probar, mientras que en la Fig. 7.1 solo hay gráficos de seis configuraciones. Específicamente, no hay diseños para la configuración  $\lambda_C = 0.2$  con  $\lambda_D = 0.1$ . Esto se debe a que, numéricamente hablando, el algoritmo no pudo encontrar una matriz positiva definida para la configuración dada. La razón detrás de ello compete al significado de los parámetros  $\lambda_C = 0.2$  y  $\lambda_D = 0.1$ . En el capítulo Capítulo 4 se tocó el hecho de si los valores de ambos parámetros eran iguales entonces se recuperaba el modelo MNL. Matemáticamente hablando,  $0.1 \neq 0.2$  sin embargo, el hecho de permitir poca variabilidad entre las opciones ocasiona que las combinaciones viables terminen por orillar el modelo con jerarquía a uno que no la posee (Fig. 7.2).

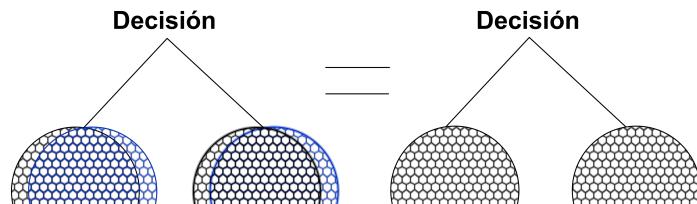


Figura 7.2: Diagrama que ilustra que poca variabilidad en un modelo con jerarquía puede convertir a éste, numéricamente hablando, en un modelo sin jerarquía y por lo tanto volviendo imposible la tarea de estimar un experimento para el modelo con jerarquía.

## 7.2 Compración vía modelo

En la Tabla 7.2 se puede observar que siempre hay un nido muy probable y un nido muy improbable. Sin embargo, el nido muy probable no es determinado por únicamente por la utilidad, sino por el producto de la disimilitud ( $\lambda_i$ 's) y la utilidad del mismo ( $V_{i,m}$ ).

---

<sup>1</sup>Solo hay una equivalencia de 5 atributos correspondiente a conjunto de elección 3.

Por ejemplo, en el caso de la configuración  $\beta = \mathbf{1}_6$ ,  $\lambda_C = 0.8$  y  $\lambda_D = 0.1$  (en la misma Tabla 7.2), el nido C y el nido D en el conjunto de elección 2 presentan una utilidad de 12.31 y 20, respectivamente; pero el nido más probable en el modelo es el C, esto se debe a que  $12.13 \times 0.8 = 9.848 > 2 = 20 \times 0.1$ . En otras palabras

En presencia de nidos con disimilitud parecida (pero nunca igual), desde el punto de vista de sus coeficientes de disimilitud, entonces el nido más probable es determinado, en muchos casos, únicamente por utilidad. En el caso de que los atributos sean positivos, entonces el nido es elegido si es el más útil; mientras que si los atributos son negativos, el nido se escoge si es el menos inútil (el más útil desde una perspectiva de pérdida). Esta observación nos permite afirmar que cuando la utilidad de un nido es negativa, se está hablando de un grado de inutilidad y no de un grado utilidad. De esta manera, el signo de  $\beta$  nos hace crear alternativas bajo dos perspectivas diferentes; en la primera, cuando el signo es positivo, se diseña bajo la búsqueda de alternativas útiles. En la segunda, cuando el signo es negativo, se diseña con el objetivo de encontrar las alternativas menos inútiles.

Si se pregunta porque no existen nidos con probabilidades similares, se debe recordar que para que exista una jerarquía la probabilidad de elección de un nido debe ser distinta a la de otro. En caso contrario, se el modelo NMNL es equivalente al MNL, donde suponemos que todas las alternativas son igualmente atractivas. Mientras la razón de que sean tan diferentes y no medianamente distintas radica, intuitivamente, en que es más sencillo observar diferencias en situaciones extremas.

## 7.3 Más allá de los diseños locales

A lo largo de estas dos primeras partes se trazó el camino a seguir para la estimación de diseños locales. Sin embargo, ¿qué ocurre con aquéllos diseños donde se propone una configuración a priori para  $\beta$ ?

La respuesta al cuestionamiento anterior, desde el punto de vista de la optimización del proceso vía evolución diferencial, es posible responderla. Ya que la implementación de un factor nuevo que modifique la función de ajuste no afecta en nada la convergencia del algoritmo, salvo en el tiempo de convergencia. Sin embargo, las rutinas empleadas para estimar cada uno de los diseños aquí mostrados, no contarón con una rutina implementada para tratar de métodos de integración numéricos en alta dimensión.

Por otro lado, [Fedorov \(2013\)](#) menciona que un enfoque ante la duda, es crear un diseño que sea óptimo estadísticamente para un conjunto de modelos. La idea de optimizar para más de un modelo puede ser laboriosa, ya que se deben establecer más de una función objetivo. Un problema, desde la problemática de la DE, equivalente es la estimación de un experimento T-óptimo y el cuál también fue atacado por [Paredes-García y Castaño Tostado \(2015\)](#).

A pesar del trabajo que falta por hacer, se puede concluir que la rutina de optimización va por buen camino y resulta para obtener diseños experimentales.

Tabla 7.2: Tabla de probabilidades y utilidades de los nidos bajo diferentes configuraciones.

CS	$\lambda_C = 0.2 \text{ y } \lambda_D = 0.9$				$\lambda_C = 0.8 \text{ y } \lambda_D = 0.1$				$\lambda_C = 0.8 \text{ y } \lambda_D = 0$			
	$V_{i,C}$	$V_{i,D}$	$p_{i,C}$	$p_{i,D}$	$V_{i,C}$	$V_{i,D}$	$p_{i,C}$	$p_{i,D}$	$V_{i,C}$	$V_{i,D}$	$p_{i,C}$	$p_{i,D}$
1	20.1269	12.1269	0.001	0.999	18.6931	15.1269	0.999998	0.000002	23	9.6931	0.9999	
2	16.0025	12.1269	0.0005	0.9995	12.3133	20	0.9996	0.0004	11.0486	14.0067	0.0224	
3	17.1269	12.0486	0.0006	0.9994	11.1269	20	0.999	0.001	19.0025	10.1269	0.997	
4	18.0009	10.3133	0.0034	0.9966	20.0486	14.3133	0.999998	0.000002	19.0067	12.1269	0.986	
5	19.1269	14.3133	0.0002	0.9998	13.3133	22	0.9998	0.0002	16.0181	20	0.0054	
6	16.1269	12.3133	0.004	0.996	16.6931	16.0181	0.999998	0.000002	15.0486	20.0067	0.0024	
7	19.0003	10.313	0.0041	0.9959	16.6931	16.6931	0.999999	0.000001	14.0025	19.0001	0.002	
8	19.0486	12.6931	0.0005	0.9995	12.6931	22	0.9996	0.0004	14.1269	18.0025	0.992	

CS	$\lambda_C = 0.2 \text{ y } \lambda_D = 0.9$				$\lambda_C = 0.8 \text{ y } \lambda_D = 0.1$				$\lambda_C = 0.8 \text{ y } \lambda_D = 0$			
	$V_{i,C}$	$V_{i,D}$	$p_{i,C}$	$p_{i,D}$	$V_{i,C}$	$V_{i,D}$	$p_{i,C}$	$p_{i,D}$	$V_{i,C}$	$V_{i,D}$	$p_{i,C}$	$p_{i,D}$
1	-14.3039	-12.8731	0.9998	0.0002	-11.8731	-11.3069	0.0002	0.9998	-18.8731	-8.9975	0.0001	
2	-13.6867	-11.9514	0.9997	0.0003	-15.8731	-9.8731	0.999992	0.000008	-16.6867	-9.9999	0.012	
3	-14.8731	-12.8731	0.9998	0.0002	-9.8731	-12.3069	0.0013	0.9987	-8	-10.6867	0.961	
4	-14.6867	-10.9514	0.999	0.001	-8.9514	-11.3069	0.0024	0.9976	-14.8731	-8.9933	0.0211	
5	-11.6867	-7.9514	0.992	0.008	-17.6867	-13.6867	0.999997	0.000003	-15.9819	-6	0.0001	
6	-16.8731	-11.9933	0.9994	0.0006	-10.9975	-13.3069	0.0006	0.9994	-19.8731	-7	1E-04	
7	-14.3069	-11.9514	0.9996	0.0004	-12.9819	-12.8731	0.999	0.001	-14.9933	-8	0.018	
8	-13.8731	-8.9993	0.9951	0.0049	-8.9997	-11.6867	0.0024	0.9976	-19.6867	-8	0.0001	

\*CS = Conjunto de elección



# Parte III

## Aplicación

Como parte final de este trabajo se presenta una aplicación de los modelos para la toma de decisión. Con ello se plantea establecer la regla de decisión existente a la hora de la primera compra de una bebida carbonatada desde el punto de vista de los atractivo visual.



# 8

## La elección del refresco

### 8.1 Una breve introducción a la elección del refresco

#### 8.1.1 *El problema*

El abuso en el consumo de bebidas azucaradas representa un problema de salud debido a ser factor relevante en problemas de salud ocasionados por la obesidad, como lo son la diabetes *mellitus* tipo 2, una larga lista de enfermedades cardiovasculares o la hipertensión (Excelsior, 2013b). Y al mismo tiempo representa un problema económico-social debido al gasto que debe hacerse para poder tratar estos padecimientos (Excelsior, 2013a).

En México, los problemas ocasionados por el hábito en la ingesta de refresco ha llegado al grado que México es el país con más muertes relacionadas con el refresco (Wade, 2013). Los dilemas a raíz del consumo de bebidas con bajo o nulo contenido nutricional oca-sionó que en 2014 se instaurara un impuesto en México sobre estas bebidas. Los resultados a un año de la entrada en vigor del impuesto tiene interpretaciones opuestas. Interpretaciones que aseguran que el impuesto es inútil (Espinosa, 2015) e interpretaciones que afirman que el impuesto está funcionando (Encarnación-Cruz, 2015). La conclusión de que el impuesto al refresco no está funcionando se centra en el hecho de que la ingesta calórica no ha disminuido ni la tasa de obesidad; mientras quienes claman que el valor agregado al refresco funciona parten del dato sobre la reducción en el consumo de refresco (Corona, 2015).

Siendo estrictos, ambas afirmaciones no son contrarias, porque no se esperaría que una persona quien consume una cierta cantidad de calorías la reduzca si necesita esa energía. Por otro lado, quienes aseguran que el consumo de refresco baja no hablan de cuál es el producto sustituto, el cual podría ser una bebida con propiedades nutricionales similares como lo son cierta clase de jugos. Por lo que el problema ocasionado por el refresco no termina en la disminución de su consumo sino en la redirección a un producto sustituto nutritivo, y para poder hacer eso es necesario entender el proceso en la toma de decisión de una bebida.

#### 8.1.2 *La elección de un alimento*

Para poder atacar el problema del consumo de refresco es necesario saber que es lo que hace a un alimento, y en especial una bebida, ser elegido por el consumidor. Una manera de ver la elección de un alimento es dividirlo en dos etapas; la primera compra y las compras posteriores (Grunert, 2005). La primera compra se basa en la *expectativa* del consumidor sobre el producto; mientras las compras posteriores tienen como base la experiencia. En otras

palabras, la primera compra se basa en la idealización del producto.

La expectativa de compra del consumidor es el resultado de una evaluación compleja del individuo. [Yeung y Morris \(2001\)](#) afirman que la expectativa de compra puede descomponerse en pérdidas; pérdida física la cual se traduce en los efectos negativos hacia la salud, una pérdida de desempeño que se refiere al sabor *esperado* del producto, una pérdida financiera que está relacionada con el costo total del producto, pérdida de tiempo que puede verse con el tiempo que implica preparar el alimento, una pérdida social y una pérdida psicológica. De esta manera, acorde con la RUT, una persona compraría el refresco que menos pérdida le ocasione.

Por otro lado, la experiencia del producto puede pensarse como una retroalimentación de las pérdidas definidas en el párrafo anterior. Así, siguiendo con lo expuesto en Capítulo 1, un producto debería ser elegido por segunda vez si se superó la expectativa del consumidor. Sin embargo, antes de pensar en las compras posteriores es necesario que la primera compra ocurra.

En México como ya se ha dicho, el refresco aún con la reducción en su consumo sigue siendo la bebida por excelencia. Aunque no todos los refrescos, en México existe una gran variedad de marcas de refresco y Coca-Cola es un amplio dominador del mercado mexicano de bebidas azucaradas en general ([Ramírez, 2015](#)). Esto puede significar que en la experiencia del consumidor el sabor del refresco sea a tal grado *bueno* que no importa el valor nutricional, los efectos a la salud o incluso su costo. Por lo que si se desea crear un sustituto al refresco debería ser un mismo refresco pero menos dañino o más nutritivo. Mientras el predominio de Coca-Cola se debe a que en comparación con los otros refrescos tiene el mejor sabor y una menor pérdida social y psicológica debido a su amplia campaña publicitaria.

Entonces la pregunta es clara, si se desea hacer una bebida sustituta al refresco actual, ¿cuáles son los atributos visuales que impactarían la expectativa del consumidor para que cambie de opción?

### **8.1.3 *Objetivo***

El objetivo de este último capítulo se centra en entender cuáles son los atributos que pueden ayudar a cambiar el predominio en la elección refrescos Cola y en general de los refrescos, los cuales son una primera opción en muchas personas a la hora de elegir una bebida, mediante el uso de los experimentos de elección discreta y los diseños experimentales desarrollados en los capítulos anteriores. Además de poder corroborar la existencia, de manera estadístico-matemática, de una jerarquía impuesta por la marca Coca-Cola a la hora de decidir sobre una bebida.

## **8.2 Metodología**

### **8.2.1 *El modelo***

Para alcanzar el objetivo del contraste entre un nuevo producto y un producto conocido, y la identificación de atributos clave que ayudan a romper la *experiencia repetida* del

producto ya conocido; se propone que la decisión del consumidor sigue un proceso de dos pasos (Fig. 8.1).

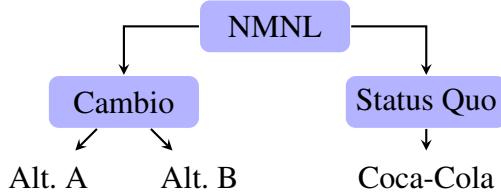


Figura 8.1: Diagrama de elección propuesto para modelar la elección de un refresco.

El proceso a dos pasos puede ser descrito por el modelo NMNL, donde la probabilidad de Cambio (C) y de *Status Quo* (SQ) están definidas como sigue

$$p_{i,C} = \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ} V_{i,SQ}}} \quad (8.1)$$

$$p_{i,SQ} = \frac{e^{\lambda_{NC} V_{i,NC}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ} V_{i,SQ}}} \quad (8.2)$$

donde

$$V_{i,C} = \log \left( \sum_{j=1}^J e^{x'_{ij} \beta + \alpha S C_j} \right) \quad (8.3)$$

$$V_{i,SQ} = \log (e^{x_{i,SQ} \beta}) = x_{i,SQ} \beta \quad (8.4)$$

de manera que la probabilidad de escoger alguna alternativa es

$$p_{i,A} = \frac{e^{x'_{iA} \beta + \alpha S C_A}}{e^{x'_{iA} \beta + \alpha S C_A} + e^{x'_{iB} \beta + \alpha S C_B}} \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ} V_{i,SQ}}} \quad (8.5)$$

$$p_{i,B} = \frac{e^{x'_{iB} \beta + \alpha S C_B}}{e^{x'_{iA} \beta + \alpha S C_A} + e^{x'_{iB} \beta + \alpha S C_B}} \frac{e^{\lambda_C V_{i,C}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ} V_{i,SQ}}} \quad (8.6)$$

$$p_{i,Coca} = \frac{e^{\lambda_{SQ} V_{i,SQ}}}{e^{\lambda_C V_{i,C}} + e^{\lambda_{SQ} V_{i,SQ}}} \quad (8.7)$$

### 8.2.2 Las alternativas

La caracterización de las alternativas fue hecha con cinco atributos ( $\mathbf{x}$ ). Los atributos fueron elegidos *a priori* porque se consideraron positivos en el aspecto visual del producto y a que son fácilmente controlables. Éstos fueron: la forma de la botella, el color de la etiqueta, el nombre del producto, el slogan del producto y la información nutrimental. En el caso de la información nutrimental, ésta fue escogida debido a que una de las características clave de un producto es su aporte en proteína y fibra. Para cada atributo se consideraron 3 niveles, bajo la restricción que las alternativas de cambio no pudieran tener el nivel correspondiente a la

Tabla 8.1: Resumen de los atributos propuestos a analizar

Atributos (x)	Nivel 1	Nivel 2	Nivel 3
Nombre	Coke	Nutrent	Nutrika
Color	Rojo/Blanco	Amarillo/Blanco	Morado/Blanco
Slogan	Destapa la felicidad	Un dulce respiro	Be free
Forma de botella	Tipo Coca-Cola	Tipo Gatorade	Tipo Ciel
Información nutrimental	Coca-Cola	Proteína	Proteína y fibra

Coca-Cola en los atributos de slogan, nombre y color de la etiqueta. Bajo estas restricciones los niveles de los atributos quedaron como muestra la Tabla 8.1.

Siguiendo la Tabla 8.1, las combinaciones a mostrar a los decisores fueron determinadas por el experimento en la Tabla 8.2, en la cual se puede observar que la positividad y la ignorancia sobre el impacto de los atributos está reflejada en  $\beta$ , la comparación con el *Status Quo* en el hecho de que la alternativa correspondiente aparece en todos los conjuntos de elección y la restricción de los atributos en que los niveles de los primeros tres atributos de las dos primeras alternativas de cada conjunto de elección nunca tienen valor 1.

### 8.2.3 La elección

Para la elección se eligieron 50 alumnos voluntarios entre las Facultades de Lenguas y Letras, Ingeniería y Psicología de la Universidad Autónoma de Querétaro con el siguiente perfil:

- Gusto por la Coca-Cola.
- No disgusto por la Guayaba.
- Actualmente cursando el grado de licenciatura.

Siguiendo con la metodología de DCEs expuesta por Mangham *et al.* (2009) a cada individuo se le pidió su decisión para cada uno de los ocho conjuntos de elección, así como tres datos personales: expediente escolar, el lugar donde comía principalmente y si practicaba deporte o no. Un ejemplo de un conjunto de elección puede observarse en la Fig. 8.2.



Figura 8.2: Ejemplo de las alternativas de un conjunto de elección

Tabla 8.2: Diseño D-óptimo local para el modelo NMNL con *Status Quo* para 8 conjuntos de elección, con 5 atributos de 2 a 3 niveles, debido que se supone que dos atributos no son exclusivos del *Status Quo* y considerando que todos los atributos afectan de manera positiva la elección y que la diferencia entre las alternativas del grupo C son diferentes con  $\lambda_C = 0.5$ .

Conjunto de elección	Diseño D-óptimo local				
	$\beta = [1, 1, 1, 1, 1]'$				
	$\lambda_C = 0.5$				
	$\lambda_{SQ} = 0$				
<b>1</b>	3	2	2	1	3
	2	3	3	2	1
	1	1	1	1	1
<b>2</b>	3	2	3	1	1
	2	3	2	1	3
	1	1	1	1	1
<b>3</b>	3	2	2	1	1
	2	2	3	1	1
	1	1	1	1	1
<b>4</b>	3	3	2	3	1
	2	2	3	1	3
	1	1	1	1	1
<b>5</b>	2	2	2	3	3
	3	3	3	1	1
	1	1	1	1	1
<b>6</b>	3	2	3	1	2
	2	3	2	2	1
	1	1	1	1	1
<b>7</b>	3	2	2	1	1
	3	3	3	1	1
	1	1	1	1	1
<b>8</b>	2	3	2	1	3
	2	2	3	3	1
	1	1	1	1	1

#### 8.2.4 Análisis estadístico

La estimación del modelo NMNL fue realizado considerando que todos los atributos como variables específicas de la alternativa y no del consumidor. La estimación fue calculada en *R* v3.2.0 junto con el paquete *mlogit* v0.2-4.

### 8.3 Resultados y discusión

Usando los parámetros estimados en Tabla 8.3 se tiene que los modelos planteados para la decisión, se pueden ver como

$$p_{iA} = \frac{e^{x'_{iA}\beta - 1.74}}{e^{x'_{iA}\beta - 1.74} + e^{x'_{iB}\beta - 1.71}} \frac{e^{0.84V_{i,C}}}{e^{0.84V_{i,C}} + 1} \quad (8.8)$$

$$p_{iB} = \frac{e^{x'_{iB}\beta - 1.71}}{e^{x'_{iA}\beta - 1.74} + e^{x'_{iB}\beta - 1.71}} \frac{e^{0.84V_{i,C}}}{e^{0.84V_{i,C}} + 1} \quad (8.9)$$

$$p_{iCoca} = \frac{1}{e^{0.84V_{i,C}} + 1} \quad (8.10)$$

donde  $\beta = [0.14, 0.43, 0.17, 0.22, 0.56]'$  y  $\lambda_C = 1.19^{-1} = 0.84$  por la forma de estimar del paquete mlogit ([Croissant, 2013](#)). Además, note que no hay intercepto ni IV estimado para el SQ, esto se debe a que es nuestro nido referencia y solo posee una alternativa. Al aplicar logaritmo se tiene que

$$\log(p_{iA}) = 0.14N + 0.43C + 0.17S + 0.22FB + 0.56IN + 0.84V_{i,C} - 1.74 - \epsilon \quad (8.11)$$

$$\log(p_{iB}) = 0.14N + 0.43C + 0.17S + 0.22FB + 0.56IN + 0.84V_{i,C} - 1.71 - \epsilon \quad (8.12)$$

$$\log(p_{iCoca}) = -\epsilon' \quad (8.13)$$

Con base al logaritmo de las probabilidades se puede decir que el supuesto de positividad de los atributos se cumple ya que todos los coeficiente asociados son positivos. De igual manera, al ser atributos que no poseen orden, se puede interpretar que aquellos con coeficientes mayores son más significativos, lo que explica porque el color y la información nutrimental posee los coeficientes más grandes. También si se observa la única diferencia entre las dos alternativas fuera del SQ es el intercepto que resulta muy parecido en magnitud, aunque la similitud de éstas queda quedada descartada por la significacia del parámetro IV que descarta la similitud, ya que de lo contrario se estaría en la presencia del modelo MNL.

Tabla 8.3: Resumen de la estimación del modelo NMNL para la toma de decisión del refresco.

*Frecuencias de las alternativas:*

	A	B	Coca-Cola	
	0.34	0.33	0.33	
<b>Coeficientes</b>	<i>Estimado</i>	<i>Error Estándar</i>	<i>Valor t</i>	<i>Pr (&gt;  t )</i>
<b>A (intercepto):</b>	-1.74	0.66	-2.64	0.008
<b>B (intercepto):</b>	-1.71	0.64	-2.65	0.008
<b>Nombre</b>	0.14	0.23	0.62	0.532
<b>Color</b>	0.43	0.19	2.31	0.021
<b>Eslogan</b>	0.17	0.20	0.86	0.390
<b>Forma de Botella</b>	0.22	0.14	1.61	0.110
<b>Información nutrimental</b>	0.56	0.15	3.75	0.000
<b>IV</b>	1.19	0.44	2.68	0.007

**Log-Verosimilitud** -415.64  
**McFadden R<sup>2</sup>** 0.054

De igual manera, en la Tabla 8.3 puede observarse la estimación resultante del experimento. En dicha tabla pueden observarse que las alternativas de cambio, así como el color e información nutrimental resultan significativas. Pero, ¿qué quiere decir esto?.

- Interceptos. En (8.8) y (8.9) los interceptos para las alternativas de cambio nos indican, además de ser significativos, que las alternativas que se propongan deben iniciar con una consideración negativa, debido a su signo. Esta información es un resultado confirmatorio de que todo producto nuevo que sale a competir en un mercado maduro inicia siendo considerado una alternativa poco atractiva respecto a los productos que ya han sido probados durante años por el consumidor, es decir, existe una cierta resistencia del consumidor hacia lo nuevo.
- Color. Los dos diseños de etiqueta propuestos fueron creados bajo un contraste sucesivo<sup>1</sup> y un contraste simultáneo<sup>2</sup>, respectivamente. Aunque todos los colores parecieran haber sido elegidos prácticamente un igual número de veces (Fig. 8.3), la estimación arroja que estadísticamente, la ligera mayor preferencia por la etiqueta morada debe ser considerada.

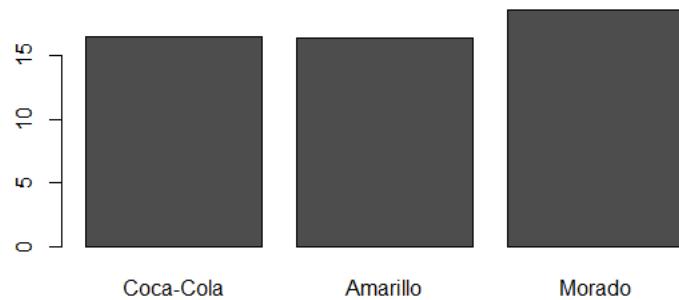


Figura 8.3: Elección relativa (el cociente de veces que fue elegida entre el número de veces que pudo ser elegida en los 8 conjunto de elección) de las opciones de acuerdo a su color.

- Información nutrimental. Si uno recuerda el ejemplo de conjunto de elección mostrado en Fig. 8.2. El hecho de que la información nutrimental sea tan visible para el consumidor difiere con el proceso de compra normal, donde el consumidor difícilmente puede consultar la información nutrimental del alimento que adquiere. Sin embargo, ¿un alimento más nutritivo es más probable de ser elegido? La respuesta es no necesariamente, en la Fig. 8.4 se puede observar que una alternativa nutritiva si es más elegida que una que no lo es, es decir, hay evidencia estadística que respalda que una bebida

<sup>1</sup> A veces también llamado complementario el cual es el contraste más usado en publicidad cuando se desea llamar la atención (Opción de morado en Fig. 8.2).

<sup>2</sup> A veces llamado puro, el cual contrasta un color cálido con otro que no lo es y permite la máxima distinción de un mensaje (Opción de amarillo en Fig. 8.2).

nutritiva es más elegible que una que no lo es; pero si dos bebidas son nutritivas, entonces depende de las características de éstas. En la Fig. 8.2 se puede observar *Proteína vs. Proteína y fibra* y proteína por sí sola resulta ser más elegida. Una posible explicación es que en la población estudiantil de las Facultades encuestadas existe un desconocimiento sobre el beneficio de la fibra, pero si se conoce que la proteína es un elemento **positivo-nutritivo**.

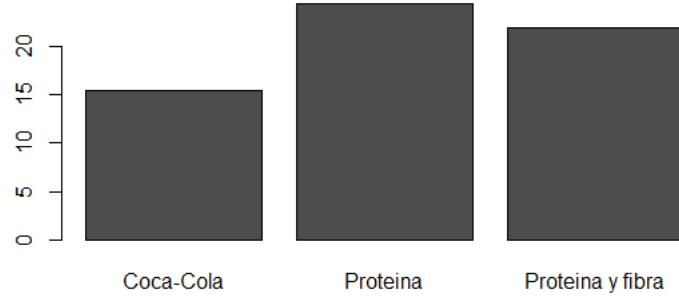


Figura 8.4: Elección relativa (el cociente de veces que fue elegida entre el número de veces que pudo ser elegida en los 8 conjunto de elección) de las opciones de acuerdo a su información nutrimental.

- IV. Que el parámetro asociado al IV haya resultado significativo, es la confirmación de que la jerarquía impuesta por el refresco de cola es estadísticamente correcta. Sobre la identificación de un paso y otro en la elección, no es posible observarlo vía las respuestas obtenidas, debido a que es imposible registrar en la respuesta de elección en esa transición. Sin embargo, la construcción del diseño mostrado en la Tabla 8.2 siempre presenta la comparación entre el *Status Quo* y las alternativas de elección propuestas, por lo cual si existiera la jerarquía de decisión debería ser observada.

Para analizar el valor del indicador de  $R^2$  de McFadden, es necesario definirlo. El indicador  $R^2$  de McFadden<sup>3</sup> es una transformación del coeficiente de correlación múltiple definido por

$$R^2 = 1 - \frac{\log L(M)}{\log L(M_0)} \quad (8.14)$$

donde  $L(M)$  es la verosimilitud del modelo con predictores y  $M_0$  es la verosimilitud del modelo sin predictores. De esta manera, un cociente pequeño entre las verosimilitudes indicaría que el modelo con predictores es mucho mejor que el modelo sin predictores, y por lo tanto, el modelo sería bueno para realizar predicciones. Sin embargo, es necesario hacer notar que a diferencia de otras aplicaciones donde se esperan valores cercanos a 1 de este coeficiente

---

<sup>3</sup>El indicador  $R^2$  es en realidad el coeficiente  $\rho^2$  debido a que el método de estimación del paquete *mlogit* es máxima verosimilitud y no mínimos cuadrados no lineales (Croissant, 2013).

que mide la bondad de ajuste, valores entre 0.4 y 0.1 se consideran excelentes (McFadden, 1977); sin embargo, el modelo estimado posee un valor de 0.05 y debe interpretarse como que el modelo estimado no posee capacidad predictiva, pero resulta **significativo**.

### 8.3.1 Análisis y resultados por grupo: la paradoja de Simpson

La paradoja de Simpson (efecto Yule-Simpson) afirma que un efecto positivo para todos los subgrupos puede ser negativo para el agregado de todos ellos (Konishi y Kitagawa, 2008). Por lo que es necesario hacer una estimación de los 3 grupos encuestados para ver si el efecto global es similar al efecto individual.

En la Tabla 8.4 se puede observar que las tendencias de los grupos de ingeniería y lenguas y letras mantienen un comportamiento similar al estimado en el modelo general. Ambos mantienen la positividad de los atributos, la poca valoración de las alternativas ajena al SQ y comparten las dos características fuertes del modelo global: la jerarquía y la relevancia de la información nutrimental. Sin embargo, psicología ofrece un contraste total, ya que los atributos presentados ni la jerarquía propuesta son parte de su proceso de toma de decisión por lo que ningún atributo es interpretable e incluso la positividad de los atributos no se mantiene.

En lo que se refiere a la significancia de los atributos en los dos primeros grupos, se puede decir que la valoración en ingeniería es menos amplia, es decir, se centran en un solo atributo, la información nutrimental, y con este hacen un juicio una vez que se tomo la decisión de elegir algo diferente al refresco de cola. Por otro lado, la valoración en Lenguas y Letras es sumamente amplia, ya que los atributos significativos son las mismas alternativas, el nombre, el color, la forma de la botella y la información nutrimental; aunque a diferencia de ingeniería y del estimado global el atributo más fuerte resulta ser el nombre.

Tabla 8.4: Resumen de la estimación del modelo NMNL para la toma de decisión del refresco por grupos.

	Ingeniería		Lenguas y Letras		Psicología	
Coeficientes	Estimado	Pr(> t )	Estimado	Pr(> t )	Estimado	Pr(> t )
<b>A (intercepto):</b>	-2,03	0.22	-4.27	0.00	0.02	0.98
<b>B (intercepto):</b>	-2.15	0.16	-3.82	0.01	-0.09	0.93
<b>Nombre</b>	0.01	0.99	1.23	0.01	-0.41	0.26
<b>Color</b>	0.43	0.29	0.61	0.08	0.43	0.12
<b>Eslogan</b>	0.10	0.81	0.34	0.43	0.17	0.57
<b>Forma de Botella</b>	0.15	0.62	0.51	0.06	0.17	0.40
<b>Info. nutrimental</b>	0.96	0.01	0.89	0.00	0.25	0.24
<b>IV</b>	0.83	0.06	0.72	0.09	1.10	0.22

Al analizar los grupos por separados se puede decir que la conclusión general de que el modelo resulta significativo no puede ser aplicado a los tres grupos ya que psicología resulta ser un enigma para el modelo propuesto.

## 8.4 Conclusión: los marcos de elección

Si se observa el conjunto de elección mostrado en la Fig. 8.2, una diferencia existente en el proceso de elección de éste es la muestra explícita de la información nutrimental. Este cambio en el proceso de elección difiere del proceso de elección tradicional donde el consumidor no tiene visible la información nutrimental.

Si se recuerda lo descrito en el Capítulo 1, la decisión de una persona depende del marco de elección. La elección entre un refresco u otro sin un marco de referencia amplio, convierte a la decisión en una elección susceptible a sesgos. En general, uno de los factores más influyentes en ese aspecto es la publicidad, la cual genera una idealización compleja del producto, haciendo que junto al hecho que su sabor es el más aceptado por el consumidor expliquen el liderazgo en ventas de Coca-Cola.

La introducción de un marco más amplio por medio de la información nutrimental explícitamente visible, transforman la decisión basada en la idealización en una elección más racional. Esta racionalización permite que exista una comparación basada en la nutrición y la idealización. De manera que un refresco como la Coca-Cola se transforma de un refresco *bueno* a un refresco *bueno* pero poco nutritivo.

Por lo tanto, para concluir, una recomendación a tomar es la creación de productos nutritivos que aprovechen la información nutrimental en la parte delantera de sus presentaciones para que, junto con el impuesto al refresco puedan convertirse en un producto sustituto nutritivo que eventualmente ayude a combatir los problemas que hoy se viven por la ingesta de refresco en México.

# A

## Algortimos para la obtención de diseños

### a.1 Algortimos y funciones base

```
1 %Especificación del experimento
2 %Sección de construcción
3 j = 4;
4 CS = 8;
5 atr = 6;
6 %Sección parametal para óptimos locales
7 beta = [1,1,1,1,1,1];
8 beta = beta';
9 lambdaC = 0.2;
10 lambdaD = 0.9;
11 %Generacion del diseño base
12 SQ = [1,1,1,1,1,1]';
13 xij = zeros(atr,j,CS);
14 %Matriz de información del diseño base
15 %En este caso se crea un diseño un 4 alternativas donde se piensa que el
16 %atributo 1 establece una jerarquía: niveles 1-2 forman una primera
17 %ramificación y los niveles 3-4 determinan una segunda ramificación.
18 for i=1:CS
19     aux = randi([1,2],1,2);
20     aux = [aux, randi([3,4],1,2)];
21     aux = [aux;randi([1,4],5,4)];
22     xij(:,:,i) = aux;
23 end
24 %Función que calcula la matriz de información que depende de
25 %beta, los conjuntos de elección y lambda, adicionalmente recibe
26 %los parámetros de control como número de alternativas (j), intervalo de
27 %conjunto de elección (1, CS), número de atributos (atr) y un parámetro
28 %de control de ramificación que en este caso es 2.
29 mi = MINMNL(beta,xij,j,1,CS,atr,lambdaC,lambdaD,2);
30 D = pinv(mi);
31 %Ciclo de búsqueda profunda - Estima 100 diseños para seleccionar el
32 %mejor.
33 for k=1:100
34     %Parámetros de la evolución diferencial
35     NP = 4; %Número de agentes
36     F = 0.8; %Paramentro de evolución.
37     agnts = zeros(atr,j,NP);
```

```

%Inicialización
39 for i=1:NP
    aux = randi([1,2],1,2);
41    aux = [aux, randi([3,4],1,2)];
42    aux = [aux;randi([1,4],5,4)];
43    agnts(:,:,:i) = aux;
end
45 %Optimización a 30 iteraciones
46 for a=1:30
47     %Cálculo del ajuste de los diseños.
48     fitG = fitness(beta,agnts,j,atr,lambdaC,lambdaD,D,NP,2);
49     [auxa,auxb] = min(fitG);
50     %Mutación de los agentes
51     [X,V] = mutation(agnts,auxb,F,NP);
52     %Corrección de los agentes
53     VC = correction(V,NP);
54     %Cálculo de ajuste de los diseños mutados.
55     fitM = fitness(beta,VC,j,atr,lambdaC,lambdaD,D,NP,2);
56     fit = [fitG,fitM];
57     [m,n] = sort(fit);
58     agntsaux = agnts;
59     %Selección de los mejores agentes (mutados y originales)
60     for i=1:NP
61         if (n(i) < 5)
62             agnts(:,:,:i) = agntsaux(:,:,n(i));
63         else
64             agnts(:,:,:i) = VC(:,:,n(i)-4);
65         end
66     end
67     t(a) = min(fit);
end
68 %Aquí termina el proceso de optimización, esto corresponde a un
69 %proceso de búsqueda profunda, el cuál consiste en comparar los
70 %los 100 diseños generados con el fin de seleccionar el mejor
71 %entre ellos.
72 xijaux = xij;
73 for i=1:(CS+1)
74     if (i > 1)
75         xij(:,:,:i-1) = agnts(:,:,:1);
76     end
77     mut = MINMNL(beta,xij,j,1,CS,atr,lambdaC,lambdaD,2);
78     deter(i) = det(mut);
79     xij = xijaux;
80 end
81 [auxaa,auxbb] = max(deter);
82 if (auxbb > 1)
83     xij(:,:,:(auxbb-1)) = agnts(:,:,:1);
84 end
85 finaldet(k) = det(MINMNL(beta,xij,j,1,CS,atr,lambdaC,lambdaD,2));
86
end

```

Algoritmo A.1: Algoritmo base de generación de diseños para el modelo NMNL

Aunque la mayoría de las acciones que se realizan en A.1 están hechas para que con un cambio único el algoritmo funcione; hay tres elementos de éste que requieren de ajustes

más profundos de manera manual:

1. La generación del diseño base así como la inicialización de los agentes. Por lo que la asignación a la variable *aux* debe modificarse para diseños con distintas características.
2. Número de iteraciones es necesario fijarlo de manera manual o hacer los cambios requeridos para que usando un ciclo *do-while* se establezca una condición de paro que dependa de la mejora generacional.
3. La selección está configurada para funcionar con 4 agentes, por lo que es necesario hacer ajustes en caso de desear más.

Por otro lado, un factor a considerar es que para la optimización se emplea la Pseudo-inversa de Moore-Penrose que es una generalización de la matriz inversa con el fin de evitar errores para matrices con determinantes con valores cercanos a una singularidad

```

1 function [ m ] = MINMNL(beta,xij,j,CSB,CSE,atr,lambdaC,lambdaD,Check )
% Cálculo de la matriz de información.
2 %beta -- Parámetro local del diseño.
3 %xij -- Diseño
4 %CSB -- Conjunto de elección inicial
5 %CSE -- Conjunto de elección final
6 %atr -- No. de atributos
7 %lambdaC y lambdaD -- Parámetros de disimilitud de los nodos
8 %Check -- Condición de la jerarquía

9
10
11 for i=CSB:CSE
12     pijD = 0;
13     pijC = 0;
14     auxC = 0;
15     auxD = 0;
16     g=1;
17     h=1;
18     part91 = 0;
19     for k=1:j
20         if ((xij(1,k,i)') > Check)
21             auxC(h) = exp(xij(:,k,i)'*beta);
22             pijC = pijC + auxC(h);
23             h = h + 1;
24         end
25         if ((xij(1,k,i)') <= Check)
26             auxD(g) = exp(xij(:,k,i)'*beta);
27             pijD = pijD + auxD(g);
28             g = g + 1;
29         end
30     end
31     ViC = log(pijC);
32     ViD = log(pijD);
33     piC = exp(lambdaC*ViC) / (exp(lambdaC*ViC)+exp(lambdaD*ViD));
34     piD = exp(lambdaD*ViD) / (exp(lambdaC*ViC)+exp(lambdaD*ViD));
35     pij_C = 0;
36     pij_D = 0;
37     pij_C = piC*(auxC/pijC);
38     pij_D = piD*(auxD/pijD);

```

```

39    h = 1;
40    g = 1;
41    part91 = 0;
42    part92 = 0;
43    part93 = 0;
44    part94 = 0;
45    part95 = 0;
46    for k=1:j
47        if ((xij(1,k,i)') > Check)
48            part91 = part91 + pij_C(h)*(xij(:,k,i)*xij(:,k,i)');
49            part92 = part92 + pij_C(h)*xij(:,k,i);
50            h = h + 1;
51        else
52            part91 = part91 + pij_D(g)*(xij(:,k,i)*xij(:,k,i)');
53            part93 = part93 + pij_D(g)*xij(:,k,i);
54            g = g + 1;
55        end
56    end
57    if (part92 == 0)
58        part92 = zeros(atr,1);
59    end
60    if (part93 == 0)
61        part93 = zeros(atr,1);
62    end
63    part92f = ((1/piC)*(lambdaC^2*(1-piC)-1))*part92*(part92');
64    part93f = ((1/piD)*(lambdaD^2*(1-piD)-1))*part93*(part93');
65    part94f = -(lambdaC*lambdaD)*part92*(part93');
66    part95f = -(lambdaC*lambdaD)*part93*(part92');
67    part96 = part91 + part92f + part93f + part94f + part95f;
68    part98 = (1-piC)*lambdaC*ViC*part92;
69    part100 = (1-piD)*lambdaD*ViD*part93;
70    part101 = piC*piD*(ViC^2);
71    part102 = piC*piD*(ViD^2);
72    mi(:,:,i) = [part96,part98,part100;part98',part101,0;part100',0,
73    part102];
74    if i==CSB
75        m = mi(:,:,i);
76    else
77        m = m + mi(:,:,i);
78    end
79 end
end

```

#### Algoritmo A.2: Cálculo de la matriz de información

La función presentada en A.2 es el cálculo de la matriz de información con base en (4.128). Además posee las condiciones de ajuste para las variables *part92* y *part93* con el fin de evitar errores para el caso de *Status Quo*. En lo que respecta al resto de la formulas empleadas, éstas corresponden a las desarrolladas en el Capítulo 4.

```

1 function [ fit ] = fitness(beta,agnts,j,atr,lambdaC,lambdaD,D,NP,mm)
2 %Función de ajuste
3 %beta -- Parámetro local del diseños.
4 %agnts -- Población inicial de diseños
5 %j -- No. de alternativas por conjunto de elección.

```

```

6 %atr -- No. de atributos
7 %lambdaC y lambdaD -- Parámetros de disimilitud de los nodos
8 %D -- Inversa de la matriz de información del diseño base
9 %NP -- No. de agentes
10 %mm -- Parámetro de optimización: Maximización(1), Minimización (Otro
11 %valor)
12
13 fit = zeros(1,NP);
14 for i=1:NP
15     %Matriz de información de cada agente
16     aux = MINMNL(beta,agnts(:,:,:,i),j,1,1,atr,lambdaC,lambdaD,1);
17     %Ajuste de acuerdo al criterio D.
18     fit(i) = trace(D*aux);
19 end
20 if (mm == 1)
21     fit = -1*fit;
22 end

```

Algoritmo A.3: Función de ajuste para la versión híbrida entre el algoritmo de intercambio de Fedorov y DE

```

1 function [ X, V ] = mutation(agnts,b,F,NP )
2 %Función básica de mutación
3 %agnts -- Diseños de la población.
4 %b -- No. de agente a mutar.
5 %F -- Parámetro de mutación.
6 %NP -- Número de agentes
7
8 X = agnts;
9 for i=1:NP
10    n = 1:NP;
11    n(i) = [];
12    n = randsample(n,3);
13    X(:,:,:,:,i) = agnts(:,:,:,:,n(1)) + F*(agnts(:,:,:,:,n(2))-agnts(:,:,:,:,n(1))) +
14    F*(agnts(:,:,:,:,n(3))-agnts(:,:,:,:,n(2)));
15 end
16 V = fix(X);
17 end

```

Algoritmo A.4: Función de mutación simple

La función presentada en A.4 presenta la versión más simple de una función de mutación, no hay combinación convexa o búsqueda local. Sin embargo, es posible obtener diseños con ella, aunque el número de iteraciones aumenta en comparación con una función de mutación más sofisticada.

```

1 function [ V ] = correction( V, NP )
2 %Función básica de corrección para los agentes de la evolución
3 %Recibe un diseño (V) y un número de agentes (NP)
4
5 for i=1:NP
6     %Corrección bloque 1
7     aux1 = V(1:1,1:2,i);
8

```

```

10 [m,n] = size(aux1);
aux1 = randi([1,2],m,n);
aux2 = (V(1:1,1:2,i) < 3-0.001 & V(1:1,1:2,i) > 0.001);
V(1:1,1:2,i) = aux2.*V(1:1,1:2,i) - (aux2-1).*aux1;
%Corrección bloque 2
14 aux1 = V(1:1,3:4,i);
[m,n] = size(aux1);
16 aux1 = randi([3,4],m,n);
aux2 = (V(1:1,3:4,i) < 5-0.001 & V(1:1,3:4,i) > 2+0.001);
V(1:1,3:4,i) = aux2.*V(1:1,3:4,i) - (aux2-1).*aux1;
%Corrección bloque 3
20 aux3 = V(2:6,1:4,i);
[m,n] = size(aux3);
22 aux3 = randi([1,4],m,n);
aux4 = (V(2:6,1:4,i) < 5-0.001 & V(2:6,1:4,i) > 0.001);
V(2:6,1:4,i) = aux4.*V(2:6,1:4,i) - (aux4-1).*aux3;
24
end
26 end

```

Algoritmo A.5: Función de mutación simple

En A.5 se presenta una función de corrección muy sencilla debido a que solo aplica una corrección en donde hay errores sin aplicar penalización. Aunque la aplicación de la corrección simple es suficiente se pueden aplicar mejoras. Por otro lado, la función presentada es sumamente manual por lo que los bloques de corrección deben ser modificados según el diseño de interés.

## a.2 Funciones opcionales

## a.3 Funciones usadas para la comparación

```

1 %Matriz de información del Logit Multinomial-
2
3 %Especificación del diseño.
j=2;
5 CS = 8;
xij = zeros(6,2,CS);
7 xij(:,:,1) = [1,2;2,2;2,1;1,2;2,1;2,2];
xij(:,:,2) = [2,2;1,2;1,2;2,2;2,1;2,1];
9 xij(:,:,3) = [2,1;2,1;1,2;1,2;1,2;2,1];
xij(:,:,4) = [2,1;2,1;1,2;1,2;2,1;1,2];
11 xij(:,:,5) = [1,2;2,1;1,2;2,1;2,1;1,2];
xij(:,:,6) = [1,2;2,1;1,2;2,1;1,2;2,1];
13 xij(:,:,7) = [2,1;1,2;1,2;2,1;2,2;1,2];
xij(:,:,8) = [2,1;1,2;2,2;2,1;2,1;1,2];
15
beta = [-1,-1,-1,-1,-1,-1];
17 beta = beta';
pij = 0;
19

```

```

% Cálculo de la matriz de información con base en la expresión exhibida
    por
21 %Goos et al. (2010)
for i=1:CS
    for k=1:j
        aux(k) = exp(xij(:,k,i)'*beta);
        pij = pij + aux(k);
    end
    aux = aux/pij;
    mi(:,:,i) = (xij(:,:,i)*(diag(aux)-aux'*aux)*xij(:,:,i)');
    if (i == 1)
        m = mi(:,:,1);
    else
        m = m + mi(:,:,i);
    end
    pij = 0;
end

```

Algoritmo A.6: Script para el cálculo de la matriz de información del modelo MNL.

```

1 %Matriz de información del Nested Logit Multinomial para la no elección
2 %con base en la expresión desarrollada por Goos et al (2010).
3
4 %Especificación del diseño.
5 j=2;
6 CS = 8;
7 xij = zeros(6,2,CS);

9 %Diseño para el modelo MNL
10 %xij(:,:,1) = [1,2;1,2;1,2;1,2;1,2;1,2];
11 %xij(:,:,2) = [2,1;1,2;1,2;1,2;2,1;2,1];
12 %xij(:,:,3) = [1,2;2,1;1,2;1,2;2,1;2,1];
13 %xij(:,:,4) = [2,1;2,1;1,2;1,2;1,2;1,2];
14 %xij(:,:,5) = [1,2;1,2;2,1;1,2;2,1;1,2];
15 %xij(:,:,6) = [2,1;1,2;2,1;1,2;1,2;2,1];
16 %xij(:,:,7) = [1,2;2,1;2,1;1,2;1,2;2,1];
17 %xij(:,:,8) = [2,1;2,1;2,1;1,2;2,1;1,2];

19 %Diseño estimado por Goos et al. (2010) para el modelo NMNL.
20 xij(:,:,:1) = [1,2;2,2;2,1;1,2;2,1;2,2];
21 xij(:,:,:2) = [2,2;1,2;1,2;2,2;2,1;2,1];
22 xij(:,:,:3) = [2,1;2,1;1,2;1,2;1,2;2,1];
23 xij(:,:,:4) = [2,1;2,1;1,2;1,2;2,1;1,2];
24 xij(:,:,:5) = [1,2;2,1;1,2;2,1;2,1;1,2];
25 xij(:,:,:6) = [1,2;2,1;1,2;2,1;1,2;2,1];
26 xij(:,:,:7) = [2,1;1,2;1,2;2,1;2,2;1,2];
27 xij(:,:,:8) = [2,1;1,2;2,2;2,1;2,1;1,2];

29 %Diseño obtenido para el modelo NMNL usando DE.
30 %xij(:,:,:1) = [2,2;2,2;3,2;2,3;2,2;2,3];
31 %xij(:,:,:2) = [2,3;2,3;2,2;2,2;2,2;3,2];
32 %xij(:,:,:3) = [2,2;2,3;3,2;2,3;2,2;2,2];
33 %xij(:,:,:4) = [2,2;2,2;2,2;2,3;3,2;2,2];
34 %xij(:,:,:5) = [2,2;3,2;2,2;2,3;2,3;3,2];
35 %xij(:,:,:6) = [2,2;3,2;2,3;2,2;2,3;2,3];

```

```

37 %xij(:,:,7) = [2,3;3,2;3,2;2,2;2,2;2,2];
38 %xij(:,:,8) = [2,3;2,2;2,3;2,2;3,2;2,2];
39 %Operación de desplazamiento en el caso del diseño obtenido mediante DE.
40 %xij = xij -1;
41
42 j=2;
43 CS = 8;
44
45 beta = [1,1,1,1,1,1];
46 beta = -beta';
47 lambdaC = 0.5;
48 pij = 0;
49
50 %Desarrollo de la expresión de Goos et al. (2010) para 8 conjunto de
51 %selección.
52 for i=1:8
53     aux = 0;
54     for k=1:j
55         aux(k) = exp(xij(:,:,k,i)'*beta);
56         pij = pij + aux(k);
57     end
58     ViC = log(pij);
59     ViNC = 0;
60     piC = exp(lambdaC*ViC)/(exp(lambdaC*ViC)+1);
61     piNC = 1/(exp(lambdaC*ViC)+1);
62     aux = piC*(aux/pij);
63     ma(:,:,i) = (xij(:,:,i))*diag(aux)+(1/piC)*(lambdaC^2 * piNC - 1)*(aux'*aux)*xij(:,:,i)');
64     me(:,:,i) = piNC*lambdaC*ViC*aux*xij(:,:,i)';
65     mo(:,:,i) = piNC*lambdaC*ViC*xij(:,:,i)*aux';
66     mu(:,:,i) = piC*piNC*(ViC^2);
67     mi = [ma(:,:,i),mo(:,:,i);me(:,:,i),mu(:,:,i)];
68     if (i == 1)
69         m = mi;
70     else
71         m = m + mi;
72     end
73     pij = 0;
74 end

```

Algoritmo A.7: Script para el cálculo de la matriz de información usando la expresión de Goos *et al.* (2010).

# B

## Diseños óptimos locales

En el Capítulo 7 se muestra una análisis de los diseños obtenidos, en este apéndice se muestran los 6 diseños obtenidos de manera explícita.

Tabla B.1: Diseños correspondientes al b y f de la Tabla 7.1.

Conjunto de elección	Experimento D-óptimo local $\beta = [1, 1, 1, 1, 1]$ $\lambda_C = 0.2$ y $\lambda_D = 0.9$						Conjunto de elección	Experimento D-óptimo local $\beta = -[1, 1, 1, 1, 1]$ $\lambda_C = 0.2$ y $\lambda_D = 0.9$							
	1	2	3	4	5	6		1	2	3	4	5	6	7	8
<b>1</b>	1	1	1	1	2	2	<b>1</b>	1	4	4	3	2	1		
	2	1	4	1	1	1		1	3	3	3	2	1		
	3	3	2	4	1	3		3	1	1	3	3	4		
	3	2	1	2	4	2		4	4	4	1	1	1		
<b>2</b>	1	1	3	2	1	1	<b>2</b>	2	3	2	1	2	2		
	2	2	1	1	3	1		1	1	4	3	3	3		
	4	1	1	1	4	1		4	1	2	2	2	4		
	4	2	2	4	1	4		3	3	2	4	1	1		
<b>3</b>	2	1	1	1	1	3	<b>3</b>	2	1	4	2	3	3		
	1	1	1	4	1	2		1	1	3	4	2	2		
	3	3	2	3	2	1		3	1	4	3	2	2		
	3	2	2	2	4	3		4	3	3	1	4	2		
<b>4</b>	1	3	1	1	2	4	<b>4</b>	2	1	4	1	2	4		
	2	1	1	2	1	4		1	1	2	3	2	2		
	4	3	2	3	1	3		4	3	1	3	3	1		
	3	2	1	1	2	2		3	2	4	3	1	3		
<b>5</b>	2	3	1	1	1	2	<b>5</b>	1	3	1	1	1	1		
	1	1	1	1	3	2		1	3	2	3	1	1		
	4	2	2	4	3	2		3	1	3	1	3	1		
	4	1	3	1	1	2		4	3	2	2	1	1		
<b>6</b>	2	4	3	3	2	3	<b>6</b>	1	3	2	2	2	2		
	1	2	4	3	4	4		2	3	1	4	3	4		
	3	1	4	3	1	3		4	3	4	2	1	3		
	3	2	1	2	2	4		3	4	4	2	4	2		
<b>7</b>	1	2	1	2	1	1	<b>7</b>	2	1	3	1	4	1		
	2	1	1	1	4	1		2	4	4	2	1	2		
	4	1	3	4	2	2		4	1	4	3	2	1		
	3	2	2	4	1	4		3	4	3	1	1	3		
<b>8</b>	1	1	1	2	1	4	<b>8</b>	1	2	1	1	2	2		
	2	1	2	1	2	1		2	2	2	2	4	2		
	3	1	1	2	3	2		4	2	3	3	3	1		
	4	3	3	4	4	3		4	2	4	1	2	1		

Tabla B.2: Diseños correspondientes al c y g de la Tabla 7.1.

Conjunto de elección	Experimento D-óptimo local $\beta = [1, 1, 1, 1, 1, 1]$ $\lambda_C = 0.8$ y $\lambda_D = 0.1$						Conjunto de elección	Experimento D-óptimo local $\beta = -[1, 1, 1, 1, 1, 1]$ $\lambda_C = 0.8$ y $\lambda_D = 0.1$					
	1	4	1	1	4	4		1	2	4	1	3	1
1	1	1	3	3	4	1	1	1	2	3	1	1	4
	4	3	1	3	4	3		3	2	3	2	1	1
	4	4	3	4	1	2		3	3	3	2	1	2
	2	4	4	2	4	4		1	1	1	3	1	3
2	1	1	1	4	1	1	2	1	3	3	1	3	1
	4	3	1	1	1	1		4	3	1	3	3	4
	3	1	2	1	3	2		3	2	3	2	3	3
	2	4	2	4	4	4		1	1	4	4	1	2
3	1	1	1	2	3	1	3	1	4	2	1	4	1
	3	2	1	1	3	1		4	2	2	2	1	1
	3	2	1	1	1	1		3	1	1	3	1	1
	1	1	3	3	2	3		1	1	4	1	1	4
4	1	2	2	1	4	4	4	1	4	3	1	1	2
	4	3	4	3	2	4		3	1	1	1	2	1
	3	3	3	4	3	1		3	3	2	2	1	1
	1	1	1	1	1	2		1	4	1	1	4	4
5	2	4	4	4	4	4	5	1	2	4	4	2	1
	2	3	1	1	1	4		3	3	3	2	3	4
	3	1	4	2	2	1		3	4	3	3	2	4
	2	2	2	1	2	3		2	1	1	3	3	4
6	2	3	3	3	1	4	6	1	2	3	3	3	2
	4	3	3	1	3	2		4	1	2	1	2	1
	3	4	2	4	1	2		3	4	2	2	2	4
	1	2	4	3	3	3		1	4	1	4	2	3
7	1	4	2	4	2	3	7	2	4	1	1	3	1
	4	4	2	1	1	4		3	1	1	2	4	2
	4	2	1	4	1	4		3	4	3	1	4	2
	1	1	1	1	1	2		1	2	2	1	2	4
8	2	4	4	4	4	4	8	1	4	1	3	3	1
	4	1	2	2	1	2		3	4	2	4	1	3
	3	2	2	1	2	2		3	2	1	1	1	1

Tabla B.3: Diseños correspondientes al d y h de la Tabla 7.1.

Conjunto de elección	Experimento D-óptimo local						Conjunto de elección	Experimento D-óptimo local						
	$\beta = [1, 1, 1, 1, 1, 1]$							$\beta = -[1, 1, 1, 1, 1, 1]$						
	$\lambda_C = 0.8$ y $\lambda_D = 0.9$							$\lambda_C = 0.8$ y $\lambda_D = 0.9$						
1	1	1	1	1	1	1	1	2	2	3	2	3	3	
	1	2	3	1	1	1		1	1	2	1	2	2	
	4	2	2	2	2	3		4	2	4	4	3	2	
	3	1	2	1	1	1		4	4	4	3	4	2	
2	2	1	1	1	1	1	2	2	1	1	1	4	1	
	1	2	2	2	1	3		1	4	4	3	3	4	
	4	4	4	3	3	4		4	3	4	2	3	2	
	4	2	4	1	3	4		4	2	3	3	4	1	
3	1	1	3	1	1	1	3	2	3	1	2	1	3	
	2	4	1	1	1	2		1	3	1	2	2	2	
	4	4	4	3	3	4		3	1	1	1	1	1	
	3	1	1	2	2	1		4	3	2	3	3	3	
4	1	1	3	1	1	1	4	1	3	3	1	3	3	
	1	2	1	1	2	2		1	1	2	1	1	3	
	3	4	4	4	4	4		3	2	4	3	2	1	
	3	4	1	1	2	1		3	1	2	4	3	4	
5	1	4	4	4	4	3	5	1	1	1	1	1	1	
	1	2	1	1	3	1		1	4	3	4	3	4	
	3	4	1	3	3	2		4	2	3	3	2	2	
	3	2	2	1	1	1		4	3	3	4	3	3	
6	1	4	3	1	1	1	6	1	4	4	3	3	3	
	1	1	2	1	2	1		2	1	1	1	1	1	
	3	1	1	1	1	2		4	2	4	4	3	3	
	4	4	4	4	3	3		4	3	4	3	4	4	
7	2	3	2	1	1	3	7	2	1	2	1	1	1	
	1	1	1	1	2	2		1	3	4	4	2	3	
	3	2	1	1	1	2		3	3	3	4	4	3	
	4	4	4	4	3	4		3	2	3	3	1	3	
8	2	4	4	4	3	3	8	1	1	2	1	2	1	
	1	1	3	1	2	1		2	4	4	3	2	3	
	3	2	2	1	2	2		4	2	3	4	3	4	
	3	1	1	1	1	1		4	1	4	4	4	4	



C

## Conjuntos de elección para la decisión del refresco

Alternativa A



Un  
dulce  
respiro

Información nutricional	
Calorías (kcal)	57
Carbohidratos (g)	13
Azúcares (g)	13
Fibra (g)	0
Grasa (g) dietética (g)	0
Proteínas (g)	0.8

Alternativa B



Be  
Free

Información nutricional	
Calorías (kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra (g)	0
Grasa (g) dietética (g)	0
Proteínas (g)	0

Alternativa C



Destapa  
la  
felicidad

Información nutricional	
Calorías (kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra (g)	0
Grasa (g) dietética (g)	0
Proteínas (g)	0

Figura C.1: Conjunto de elección 1 correspondiente al diseño de la Tabla 8.2.

Alternativa A



Información nutricional	
Calorías (Kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra dietética (g)	0
Grasa (g)	0
Proteínas (g)	0

Alternativa B



Información nutricional	
Calorías (Kcal)	57
Carbohidratos (g)	13
Azúcares (g)	13
Fibra dietética (g)	0
Grasa (g)	0
Proteínas (g)	0.8

Alternativa C



Información nutricional	
Calorías (Kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra dietética (g)	0
Grasa (g)	0
Proteínas (g)	0

Figura C.2: Conjunto de elección 2 correspondiente al diseño de la Tabla 8.2.

Alternativa A



Un  
dulce  
respiro

Información nutricional	
Calorías (kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra dietética (g)	0
Graje (g)	0
Proteínas (g)	0

Alternativa B



Be  
free

Información nutricional	
Calorías (kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra dietética (g)	0
Graje (g)	0
Proteínas (g)	0

Alternativa C

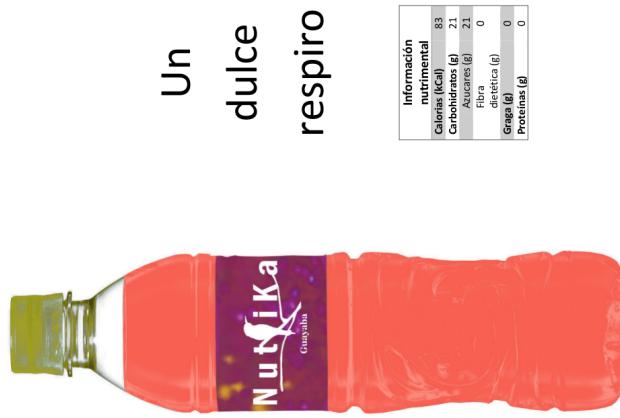


Destapa  
la  
felicidad

Información nutricional	
Calorías (kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra dietética (g)	0
Graje (g)	0
Proteínas (g)	0

Figura C.3: Conjunto de elección 3 correspondiente al diseño de la Tabla 8.2.

Alternativa A



Información nutricional	
Calorías (Kcal)	83
Carbohidratos (g)	57
Azúcares (g)	21
Fibra dietética (g)	0
Grasa (g)	0
Proteínas (g)	0

Alternativa B



Información nutricional	
Calorías (Kcal)	57
Carbohidratos (g)	13
Azúcares (g)	13
Fibra dietética (g)	0
Grasa (g)	0
Proteínas (g)	0.8

Alternativa C



Información nutricional	
Calorías (Kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra dietética (g)	0
Grasa (g)	0
Proteínas (g)	0

Figura C.4: Conjunto de elección 4 correspondiente al diseño de la Tabla 8.2.

Alternativa A



Alternativa B



Alternativa C



Figura C.5: Conjunto de elección 5 correspondiente al diseño de la Tabla 8.2.

Alternativa A



Información nutricional	
Calorías (Kcal)	47
Carbohidratos (g)	10
Azúcares (g)	8
Fibra (g)	2
dietética (g)	0
Grasa (g)	0
Proteínas (g)	0,8

Alternativa B



Información nutricional	
Calorías (Kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra (g)	0
dietética (g)	0
Grasa (g)	0
Proteínas (g)	0

Alternativa C



Información nutricional	
Calorías (Kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra (g)	0
dietética (g)	0
Grasa (g)	0
Proteínas (g)	0

Figura C.6: Conjunto de elección 6 correspondiente al diseño de la Tabla 8.2.

Alternativa A



Un  
dulce  
respiro

Información nutrimental	
Calorías (kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra	0
Azúcares dietética (g)	0
Griegos (g)	0
Proteínas (g)	0

Alternativa B



Be  
free

Información nutrimental	
Calorías (kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra	0
Azúcares dietética (g)	0
Griegos (g)	0
Proteínas (g)	0

Alternativa C



Destapa  
la  
felicidad

Información nutrimental	
Calorías (kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra	0
Azúcares dietética (g)	0
Griegos (g)	0
Proteínas (g)	0

Figura C.7: Conjunto de elección 7 correspondiente al diseño de la Tabla 8.2.

Alternativa A



Información nutricional	
Calorías (Kcal)	57
Carbohidratos (g)	13
Azúcares (g)	13
Fibra dietética (g)	0
Grasa (g)	0
Proteínas (g)	0

Alternativa B



Información nutricional	
Calorías (Kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra dietética (g)	0
Grasa (g)	0
Proteínas (g)	0

Alternativa C



Información nutricional	
Calorías (Kcal)	83
Carbohidratos (g)	21
Azúcares (g)	21
Fibra dietética (g)	0
Grasa (g)	0
Proteínas (g)	0

Figura C.8: Conjunto de elección 8 correspondiente al diseño de la Tabla 8.2.



# Bibliografía

- Allenby, G. M. y Rossi, P. E. (1998). Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89(1-2):57–78.
- Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, 41(3):258.
- Atkinson, A. C., Donev, A. N., y Tobias, R. (2007). *Optimum experimental designs, with SAS*. Oxford University Press, Oxford; New York.
- Berger, J., Meredith, M., y Wheeler, S. C. (2008). Contextual priming: Where people vote affects how they vote. *Proceedings of the National Academy of Sciences*, 105(26):8846–8849.
- Burgess, L. y Street, D. J. (2005). Optimal designs for choice experiments with asymmetric attributes. *Journal of Statistical Planning and Inference*, 134(1):288–301.
- Corona, S. (2015). Los productores de refresco en México dicen que el impuesto es inútil. *El País*.
- Croissant, Y. (2013). Estimation of multinomial logit models in r: The mlogit packages. <https://cran.r-project.org/web/packages/mlogit/vignettes/mlogit.pdf>. [En linea; acceso 02-Agosto-2015].
- Demirkale, F., Donovan, D., y Street, D. J. (2013). Constructing D-optimal symmetric stated preference discrete choice experiments. *Journal of Statistical Planning and Inference*, 143(8):1380–1391.
- Dette, H. y Titoff, S. (2009). Optimal discrimination designs. *The Annals of Statistics*, 37(4):2056–2082.
- Encarnación-Cruz, L. M. (2015). El impuesto al refresco sí funciona. *El Universal*.
- Espinosa, E. (2015). Impuesto al refresco causa 8,800 despidos y no combate obesidad. *La Unión de Morelos*.
- Excelsior (2013a). Asciende a 80 mil mdp tratamiento de sobrepeso y obesidad. *Excelsior Nacional*.
- Excelsior (2013b). Preocupa obesidad y diabetes: Ahued. *Excelsior Comunidad*.

- Fedorov, V. V. (2013). *Optimal design for nonlinear response models*. Chapman & Hall/CRC biostatistics series. Taylor & Francis, Boca Raton.
- Feoktistov, V. (2006). *Differential evolution: in search of solutions*. Number v. 5 in Optimization and its applications. Springer Science+Business Media, New York.
- Gilbert, D. T., Krull, D. S., y Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of personality and social psychology*, 59(4):601.
- Goos, P., Vermeulen, B., y Vandebroek, M. (2010). D-optimal conjoint choice designs with no-choice options for a nested logit model. *Journal of Statistical Planning and Inference*, 140(4):851–861.
- Greene, W. H. y Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8):681–698.
- Grunert, K. G. (2005). Food quality and safety: consumer perception and demand. *European Review of Agricultural Economics*, 32(3):369–391.
- Hardie, B. G. S., Johnson, E. J., y Fader, P. S. (1993). Modeling Loss Aversion and Reference Dependence Effects on Brand Choice. *Marketing Science*, 12(4):378–394.
- Jafari, H. (2013). Locally D-Optimal Design for a Logit Model in Discrete Choice Experiment. *Communications in Statistics - Theory and Methods*, 42(5):867–882.
- Kahneman, D. (2012). *Pensar rápido, pensar despacio*. Debate.
- Kahneman, D. y Tversky, A. (1984). Choices, values, and frames. *American psychologist*, 39(4):341.
- Kessels, R., Jones, B., Goos, P., y Vandebroek, M. L. (2006). An Efficient Algorithm for Constructing Bayesian Optimal Choice Designs. *SSRN Electronic Journal*.
- Konishi, S. y Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. Springer Series in Statistics. Springer.
- Li, H. y Zhang, L. (2014). A discrete hybrid differential evolution algorithm for solving integer programming problems. *Engineering Optimization*, 46(9):1238–1268.
- Louviere, J., Islam, T., Wasi, N., Street, D., y Burgess, L. (2008). Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? *Journal of Consumer Research*, 35(2):360–375.
- Louviere, J. J. (2006). What You Don't Know Might Hurt You: Some Unresolved Issues in the Design and Analysis of Discrete Choice Experiments. *Environmental and Resource Economics*, 34(1):173–188.

- Louviere, J. J., Hensher, D. A., y , Joffre Dan, S. (2000). *Stated choice methods: analysis and applications*. Cambridge University Press, Cambridge, UK; New York, NY, USA.
- Louviere, J. J. y Lanksar, E. (2009). Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Economics, Policy and Law*, 4(04):527.
- Mangham, L. J., Hanson, K., y McPake, B. (2009). How to do (or not to do) ... Designing a discrete choice experiment for application in a low-income country. *Health Policy and Planning*, 24(2):151–158.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.
- McFadden, D. (1977). Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments. Technical report.
- Morewedge, C. K. y Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in cognitive sciences*, 14(10):435–440.
- Paredes-García, W. J. y Castaño Tostado, E. (2015). Differential evolutionary algorithms in the building process of optimal experimental designs.
- Price, K. V., Storn, R. M., y Lampinen, J. A. (2005). *Differential evolution a practical approach to global optimization*. Springer, Berlin; New York.
- Pronzato, L. y Pázman, A. (2013). *Design of experiments in nonlinear models: asymptotic normality, optimality criteria and small-sample properties*. Number v. 212 in Lecture notes in statistics. Springer, New York.
- Pukelsheim, F. (2006). *Optimal design of experiments*. Number 50 in Classics in applied mathematics. SIAM/Society for Industrial and Applied Mathematics, Philadelphia, classic ed edition.
- Rabin, M. et al. (2000). Risk aversion and expected-utility theory: A calibration theorem. *University of California at Berkeley, Department of Economics*.
- Ramírez, M. (2015). *Coca cola es la marca favorita de los mexicanos*. Milenio.
- Revelt, D. y Train, K. (1998). *Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level*. Review of Economics and Statistics, 80(4):647–657.
- Ryan, M. y Skatun, D. (2004). *Modelling non-demanders in choice experiments*. *Health Economics*, 13(4):397–402.
- SAS (2013). *SAS/QC(R) 9.2 User's Guide, Second Edition*. [http://support.sas.com/documentation/cdl/en/qcug/63922/HTML/default/viewer.htm#qcug\\_optex\\_a0000000398.htm](http://support.sas.com/documentation/cdl/en/qcug/63922/HTML/default/viewer.htm#qcug_optex_a0000000398.htm). [En linea; acceso 02-Agosto-2015].

- Schneider, E. R. y Krohling, R. A. (2014). A hybrid approach using TOPSIS, Differential Evolution, and Tabu Search to find multiple solutions of constrained non-linear integer optimization problems. Knowledge-Based Systems, 62:47–56.*
- Seber, G. A. F. y Wild, C. J. (2003). Nonlinear regression. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, N.J.*
- Street, D. (2004). Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments. Journal of Statistical Planning and Inference, 118(1-2):185–199.*
- Thurstone, L. L. (1994). A law of comparative judgment. Psychological Review, 101(2):266–270.*
- Tversky, A. y Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157):1124–1131.*
- Tversky, A. y Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychological review, 90(4):293.*
- Tversky, A. y Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. The quarterly journal of economics, pages 1039–1061.*
- Wade, L. (2013). México, el país con más muertes relacionadas al consumo de refresco. CNN Vida y Salud.*
- Whittlesea, B. W., Jacoby, L. L., y Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. Journal of Memory and Language, 29(6):716–732.*
- Yeung, R. M. y Morris, J. (2001). Food safety risk: Consumer perception and purchase behaviour. British Food Journal, 103(3):170–187.*