



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

DESARROLLO DE CUESTIONARIOS EFICIENTES PARA EL MODELO DE
ELIMINACIÓN POR ASPECTOS

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTIÓN DE OPERACIONES
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

IGNACIO ALEJANDRO CORREA FALCONI

PROFESOR GUÍA:
DENIS ROLAND SAURÉ VALENZUELA

MIEMBROS DE LA COMISIÓN:
RICARDO ESTEBAN MONTOYA MOREIRA
JOSÉ ANTONIO SOTO SAN MARTÍN

Este trabajo ha sido parcialmente financiado por Fondecyt Regular 1151395.

SANTIAGO DE CHILE

2017

RESUMEN DE LA MEMORIA PARA OPTAR AL TÍTULO DE: Ingeniero Civil Matemático y grado de Magíster en Gestión de Operaciones.

POR: Ignacio Alejandro Correa Falconi.

FECHA: 9 de marzo de 2017.

PROFESOR GUÍA: Denis Roland Sauré Valenzuela.

DESARROLLO DE CUESTIONARIOS EFICIENTES PARA EL MODELO DE ELIMINACIÓN POR ASPECTOS

En esta tesis se estudia el desarrollo de cuestionarios eficientes para el modelo de elección discreta de eliminación por aspectos. El estudio está motivado por la teoría de error desarrollada por Kohli, Jedidi y Montoya para este modelo, donde se calculan de forma explícita las probabilidades de elección de dicho modelo, y por ser más general que otros modelos más populares, como el logit o el *ranking*.

A diferencia de los modelos clásicos, este modelo no cabe dentro de los llamados modelos compensatorios (que suponen que todas las características relevantes son consideradas en el proceso de elección) por lo que las técnicas ya existentes o bien deben ser adaptadas o bien descartadas por otras nuevas. El objetivo de esta tesis es resolver el problema de obtener cuestionarios —acordes al modelo de eliminación por aspectos— que permitan recuperar los parámetros intrínsecos de dicho modelo de manera eficiente.

Se modela el problema como uno de optimización en dos niveles, cuya resolución entrega cuestionarios con las características antes mencionadas. Se provee de un algoritmo computacional para buscar dichos cuestionarios para el caso de dos y tres alternativas por pregunta, además de una cota teórica en el número de preguntas.

Se obtienen resultados con respecto a la optimalidad local (D -eficiencia) de los cuestionarios. Para los casos de dos y tres alternativas por conjunto de elección, el resultado es que los cuestionarios arrojados por el programa son más eficientes a la hora de estimar los parámetros, pero no lo son si el criterio considerado es la distancia del parámetro estimado al real. Además, se describen los cuestionarios mediante características que son relevantes en los modelos de regresión lineal, y se analiza qué tanto varían dichas características con respecto a los cuestionarios clásicos. Como resultado secundario se obtienen, para los casos antes mencionados, cotas en el número de preguntas máximas necesarias para construir un cuestionario. Estas cotas son en cierto sentido restrictivas, pero necesarias a la hora de resolver el problema de optimización. Un tercer resultado que se obtiene tiene que ver con que los tiempos de resolución del algoritmo son aceptables en las pruebas realizadas, pero el problema no es escalable para aplicar esto en instancias más grandes.

Se concluye que el procedimiento antes descrito permite obtener diseños para el modelo de eliminación por aspectos que son más eficientes que los modelos compensatorios clásicos para poder estimar los parámetros. Esto sugiere usar el modelo de eliminación por aspectos cuando existe evidencia que apoye el hecho de que los respondientes no realizan el ejercicio de considerar todas las características.

«*The harder the world, the fiercer the honour*».
—Dancer, en *Memories of Ice*, de Steven Erikson.

«*No lo pensé, sólo actué*».



Agradecimientos

Gracias infinitas a mi familia por ser ese apoyo constante e incondicional, por haberme acompañado en esta alocada decisión académica y en toda mi vida. A mi padres Elizet y Mario, y mis hermanos Sofía y Cristián por nunca faltar: ustedes me formaron y así salí, qué le van a hacer. A mis tías y tíos, primas y primos, abuelas y abuelos. Muy en particular a mis abuelos paternos Óscar y María, q. e. p. d. Se les echa de menos.

Al departamento 1007 de Stannis the Mannis, el verdadero rey de Westeros —*bend the knee or you'll be destroyed*—, por ser esa fuente de amistad, amor, conversaciones e imbecilidades que estuvo conmigo todo este tiempo. Tomás, Rodrigo, Matías, Renzo, Carolina y Valentina, todo me lo llevo en el corazón. «*Gracias por haber venido esta noche, mis amigos, hasta acá. Muchas gracias de verdad. Por los coros compartidos, los aplausos recibidos, ojalá que nos volvamos a encontrar*».

A la Tefa y a Frodo por decirme «toma CVV con Correa» y «yo creo que te tincaría el MGO».

A todo el DIM por acogerme como un bicho raro más. Muy en particular a mi generación, por estar juntos en esta ardua tarea, y a la generación un año más vieja, por invitarnos a ser parte de su círculo. Más en particular a Nico Sanhueza y JT por sus valiosos aportes a esta tesis. Su escriba ya termina una fase. No volverán a pedirme mis apuntes nunca más. A cada uno de ustedes, n gracias, $\forall n \in \mathbb{N}$.

A todas las amistades hechas en la sala 11 de CRI/Seishin, por explotar y potenciar mi ñoñeza jugando rol, juegos de mesa y sesionando capítulos de Los Hombres También Lloran.

A «La Salita», por todo el ambiente grato de trabajo, amistad, café y belicosidad. Se agradece la compañía y el sufrimiento mutuo por sacar adelante nuestros trabajos finales y primeras pegas.

A todos los académicos, profesores y funcionarios del DIM y del MGO, por su trabajo como académicos formadores. Al profesor Rafael Correa por iniciarme en esta carrera; a los profesores Raul Gouet, Servet Martínez y Daniel Remenik por generar y reafirmar mi gusto por las probabilidades; a los profesores José Verschae, Maya Stein, Roberto Cominetti, Martín Matamala y José Correa por complementar mi formación en optimización, algoritmos y matemáticas discretas; y a los profesores Patricio Felmer, Andrea Rodríguez, Jaime San Martín, Cristián Reyes, Sergio Jara, Sebastián Ríos y Andrés Weintraub por enseñarme tópicos diversos con una habilidad docente particular e insuperable.

A los profesores Denis Sauré y Ricardo Montoya, del DII y del MGO, y al profesor José Soto del DIM por ser mis profesores guía, coguía y miembro de la comisión. Su apoyo, disponibilidad, acompañamiento, sugerencias y retroalimentación fueron y son invaluable. Ha sido un agrado trabajar con ustedes, y se han convertido en un referente a seguir en esta área de la investigación académica.

Tabla de contenido

Introducción	1
1. Definiciones básicas y resultados previos	3
1.1. Los modelos de elección discreta	3
1.2. Las probabilidades de elección en los modelos de elección discretos multinomiales	9
1.3. El logit multinomial	9
1.4. El modelo EBA como modelo logit	10
1.5. Diseño de experimentos: cuestionarios eficientes	12
2. Un algoritmo de optimización para el desarrollo de cuestionarios con dos alternativas	14
2.1. Formato de las preguntas	15
2.2. Las probabilidades	16
2.3. El problema MIP binivel	16
2.4. Detalles técnicos de la resolución del problema	18
2.5. Resultados y análisis	24
3. Un algoritmo de optimización para el desarrollo de cuestionarios con tres alternativas	32
3.1. Formato de las preguntas	32
3.2. Las probabilidades	32
3.3. El problema MIP binivel	34
3.4. Detalles técnicos de la resolución del problema	34
3.5. Resultados y análisis	38
Conclusión	42
Bibliografía	45
A. Propiedades de la distribución Gumbel	48
B. Algoritmo de generación de columnas para el problema BiMIP y adaptación	49
B.1. El algoritmo general	49
B.2. Adaptación al problema de la tesis	52
C. Verosimilitud y Matrices de Información del modelo EBA	53
C.1. El caso de dos alternativas	53
C.2. El caso de tres alternativas	54

D. Ejemplos de diseño	57
D.1. Dos alternativas, caso 2×3^2	57
D.2. Tres alternativas, caso $2 \times 3 \times 4$	58

Índice de ilustraciones

2.1. Cantidad de preguntas según número de atributos y niveles para el caso de dos alternativas.	22
2.2. Comparación de la DAP.	30
2.3. Comparación de la correlación.	30
2.4. Comparación del nivel de traslape.	31
3.1. Cantidad de preguntas según número de atributos y niveles para el caso de tres alternativas.	37
3.2. Comparación de la DAP.	42
3.3. Comparación de la correlación.	42
3.4. Comparación del nivel de traslape.	42

Índice de tablas

2.1.	Número de variables y tiempo de resolución para distintas instancias.	25
2.2.	Comparación entre el número de preguntas para distintas instancias.	26
2.3.	D^{EBA} -error para el caso $2 \times 3 \times 3$ versus número de preguntas.	27
2.4.	Datos de los mejores cuestionarios EBA por contexto.	27
2.5.	Pérdida de eficiencia según distintas métricas.	28
3.1.	Número de variables y tiempo de resolución para distintas instancias.	38
3.2.	Comparación entre el número de preguntas para distintas instancias.	39
3.3.	D^{EBA} -error para el caso $2 \times 3 \times 3$ versus número de preguntas.	40
3.4.	Datos de los mejores cuestionarios EBA por contexto.	40
3.5.	Pérdida de eficiencia según distintas métricas.	41

Introducción

En el contexto de la teoría de elección discreta, individuos se enfrentan a situaciones en donde deben elegir entre dos o más alternativas, y los investigadores intentan describir, explicar y predecir cómo los respondientes eligen. Para esto, la elección se modela de forma matemática, involucrando variables aleatorias, probabilidades de elección y combinatoria. Cada modelo trae sus propias especificaciones, dependiendo de cómo se desea describir el proceso de elección.

Una de las técnicas ampliamente usadas para poder resolver este problema es el análisis conjunto. Esta técnica estadística supone que las alternativas están descritas por atributos, e intenta determinar cómo la gente valora dichos atributos. Su objetivo es determinar qué combinación de un número limitado de atributos (llamado «experimento», «diseño» o «cuestionario») es más influyente en la elección. Se muestra un conjunto de potenciales alternativas y, analizando cómo se prefieren algunas a otras, la valoración implícita de los atributos que conforman a la alternativa pueden ser determinados. Dichas valoraciones pueden ser usadas para crear modelos que estimen cuotas de mercado, ingresos e incluso qué tan provechoso puede ser un producto hipotético.

Una mayoría de los diseños antes descritos suponen una valoración compensatoria. Esto se refiere a que una alternativa es valorada por todos y cada uno de los atributos que la definen, y que cada respondiente considera y sopesa todas estas características cuando debe elegir una alternativa por sobre otra. Bajo este marco, la construcción de diseños que permiten recuperar las valoraciones es conocida y abunda (por ejemplo, [HZ96, SW02, TH07]).

El trabajo realizado se enmarca en un modelo específico de elección discreta: el modelo de eliminación por aspectos (EBA) de Tversky [Tve72a, Tve72b]. Este modelo supone que el respondiente descarta secuencialmente alternativas, bajo un criterio aleatorio, hasta quedarse con una, la que es elegida. Dicho modelo no cae en la suposición de una valoración compensatoria, pues puede ocurrir que el respondiente no considere todos los atributos en juego. Una de las bases que atañe a este trabajo es el artículo de Kohli, Jedidi y Montoya [KJM14], en donde se le dota al modelo EBA de una teoría de error que permite recuperar las probabilidades de elección con una fórmula cerrada. Hasta entonces, dichas probabilidades sólo poseían una fórmula recursiva.

El objetivo principal de esta tesis se centra en estudiar el problema de construir cuestionarios eficientes que permitan recuperar los parámetros del modelo EBA. Como se explica antes, las técnicas clásicas que conciernen a los modelos compensatorios no pueden ser aplicadas en este caso pues el modelo EBA es no compensatorio.

El principal resultado teórico de este trabajo es el planteamiento de un problema de optimización mixto en dos niveles que permite obtener cuestionarios que, *a priori*, permitirían recuperar las valoraciones (o parámetros) del modelo. Este problema toma en cuenta el contexto (abstracto) de atributos y niveles que definen a las alternativas, el número de alternativas que se desea mostrar, y cuántas elecciones (preguntas) se desean plantear. Además, se proveen cotas teóricas

en este número de preguntas que aseguran, en la medida de tener suficientes datos, la estimación de los parámetros.

Los resultados numéricos tienen que ver con la optimalidad de los cuestionarios obtenidos. Es común en el área estudiar la matriz de información de Fisher del modelo, e intentar maximizar alguna función a valores reales de dicha matriz. Los criterios tradicionales de optimalidad son los invariantes de la matriz que, algebraicamente, se traduce en funcionales de los valores propios de la matriz. Por mencionar algunos, la D -eficiencia maximiza el determinante de la matriz, la A -eficiencia maximiza la traza y la E -eficiencia maximiza el valor propio más pequeño. Esto se traduce en minimizar de alguna manera la varianza del modelo, puesto que la inversa de la matriz de información es la matriz de varianzas-covarianzas asintótica. Se estudian además resultados descriptivos de los cuestionarios, que aseguran buenas propiedades en los modelos compensatorios, y se discute el porqué los cuestionarios EBA poseen o no estas características. Por último, se provee de resultados con respecto a los tiempos de resolución del algoritmo y qué tan escalable es.

En el capítulo 2 se estudia el problema considerando que cada pregunta del cuestionario posee dos alternativas. Este primer caso es más sencillo, pues las probabilidades de elección tienen una forma similar al modelo logit. Simulando respuestas, se encuentra que en este caso los cuestionarios obtenidos por el programa presentan una mejor D -eficiencia que los cuestionarios compensatorios. Asimismo, se concluye que estos cuestionarios pierden menos eficiencia a la hora de estimar los parámetros cuando el modelo subyacente es compensatorio que viceversa. Las características descriptivas antes mencionadas no se aprecian en los cuestionarios EBA, lo que es esperable pues este modelo no es lineal.

En el capítulo 3 se estudia la primera extensión natural del problema: cada pregunta posee tres alternativas. La estrategia de resolución es la misma: se plantea un problema de optimización en dos niveles para obtener cuestionarios que permitirían estimar los parámetros. Luego de resolver el problema y obtener los cuestionarios se simulan respuestas y se estiman los parámetros para medir su D -eficiencia. La conclusión es similar a la obtenida en el capítulo 2: se obtienen cuestionarios para el modelo EBA que pierden menos eficiencia a la hora de estimar los parámetros cuando el modelo subyacente es compensatorio que viceversa, y las características descriptivas antes mencionadas no se aprecian pues el modelo no es lineal.

Por último, en la Conclusión se plantea una generalización de las probabilidades de elección y, en consecuencia, del problema de optimización planteado. Se plantea una conjetura con respecto a las cotas provistas para el número de preguntas, y se discuten además otros métodos de optimización para obtener los cuestionarios, como la programación semidefinida positiva y la programación cónica de segundo orden.

Capítulo 1

Definiciones básicas y resultados previos

A continuación se entrega un breve resumen de los términos básicos con los cuales se trabaja, con el fin de homologar notación y despejar posibles dudas. En la mayoría de este capítulo se siguen las ideas del libro de Anderson, de Palma y Thisse [AdT92], adaptando la notación al artículo de Kohli, Jedidi y Montoya [KJM14], más reciente y punto inicial del presente trabajo.

1.1. Los modelos de elección discreta

El contexto de la teoría de elección discreta es simple: por un lado, consumidores (o individuos) poseen una amplia variedad de gustos y se enfrentan típicamente a situaciones de elección; por otro lado, firmas que reconocen las preferencias idiosincráticas de los consumidores, y por ende su disposición a pagar más por variantes que se ajusten mejor a sus gustos, producen bienes y compiten entre ellas, intentando capturar una fracción —lo suficientemente grande— de la demanda.

El objeto de estudio de estos modelos para poder competir por la demanda es estudiar el comportamiento de los consumidores. ¿De qué forma? Como lo explica McFadden [McF73], «los modelos de elección discreta parten de la suposición de que cada consumidor, a la hora de una decisión discreta, escoge la sola opción que entrega la utilidad más grande, mientras que desde el punto de vista de un observador externo (firmas), la utilidad se describe como una variable aleatoria que refleja diferencias no observables en los gustos». Esto implica que las firmas le atribuyen *probabilidades de compra* a los consumidores, y estas probabilidades dependen de características observables (por ejemplo, precio, locación, calidad) así como de propiedades de la distribución del gusto. Las características observables pueden ser usadas por las firmas para manipular a su ventaja las probabilidades de elección. Éstas son las variables estratégicas que las firmas despliegan para competir por consumidores.

La «utilidad aleatoria» no significa que el comportamiento del consumidor sea necesariamente estocástico. Cada consumidor puede ser perfectamente racional en su elección, pero las firmas no pueden observar y medir todos los factores que determinan las preferencias de distintas variantes. Lo mejor que pueden hacer las firmas es representar estas reglas de decisión de consumo construyendo probabilidades de elección, aunque los consumidores escojan de manera determinista.

Más en concreto, considérese un individuo que debe escoger una sola alternativa de un conjunto de elección finito de alternativas C (por ejemplo, C puede ser un conjunto de productos

diferenciados). No es imperativo que C contenga una alternativa «genuina». En efecto, se puede añadir una *alternativa exterior* que corresponda a la opción «ninguna de las anteriores» (así, por ejemplo, $C = \{\text{matemáticas, industrias, computación, ninguna de las anteriores}\}$). Cada individuo posee una relación de preferencia \succsim , definida sobre C , que cumple:

1. completitud: $a \succsim b$ o $b \succsim a$ para todo $a, b \in C$ con $a \neq b$;
2. reflexividad: $a \succsim a$ para todo $a \in C$ y
3. transitividad: $a \succsim b \wedge b \succsim c \implies a \succsim c$ para todo $a, b, c \in C$.

Como C es finito, existe una alternativa óptima $a^* \in C$ para el individuo, definida como aquella tal que $a^* \succsim a$ para todo $a \in C$. Se ignoran los casos en donde varias alternativas alcancen el máximo. Además, siempre es posible construir una función de utilidad (determinista) que represente las preferencias del individuo. Tal función $U : C \rightarrow \mathbb{R}$ ha de cumplir que

$$U(a) \geq U(b) \iff a \succsim b.$$

Así, $a^* = \arg \max\{U(a) : a \in C\}$.

El enfoque anterior ha sido criticado por muchos estudiosos (más notablemente por Tversky [Tve72a, Tve72b]), quienes postulan que la gente experimenta inconsistencia e incertidumbre cuando debe elegir. De acuerdo a esto, se supondrá que a la hora de enfrentarse a una elección entre $a \in C$ y $b \in C$, existe una probabilidad $\mathbf{P}(a, b)$ de elegir la alternativa a sobre b .

La pregunta natural que surge concierne a los factores que definen dicha probabilidad. ¿Es resultado del comportamiento intrínsecamente probabilista del consumidor, o corresponde a la inhabilidad del modelador de medir con precisión dicho comportamiento (o ambas)? Existen argumentos (como los de Quandt [Qua56] y Manski [Man77]) a favor de ambas posturas. De acuerdo a Anderson *et al.* ambos factores derivan modelos idénticos, pero para efectos de modelar el comportamiento del individuo, parece ser de poca relevancia saber si dicho comportamiento es o no intrínsecamente probabilista, pues el interés está en construir demanda.

A continuación se mencionan los modelos básicos de la teoría de elección discreta. Una interpretación a analizar es que la *regla de decisión es estocástica*, mientras que la utilidad es determinista. La otra interpretación es la contraria: la regla de decisión es determinista y la *utilidad es estocástica*. En cualquier caso, las variables básicas son las probabilidades de elección, que pueden ser especificadas de antemano.

1.1.1. Modelos con reglas de decisión estocástica

La idea de estos modelos radica en el hecho de que la utilidad de las alternativas es determinista, pero el proceso de elección es aleatorio. Cada individuo tiene una probabilidad (positiva) de elegir cualquier alternativa, y no necesariamente la que reporte una mayor utilidad. Se presentan dos modelos que reflejan esta idea. El primero fue presentado por Luce [Luc59] y el segundo por Tversky [Tve72a, Tve72b], con un comentario posterior de que el modelo de Luce es un caso especial del modelo de Tversky.

En lo que sigue se supone un conjunto de elección C y subconjuntos del mismo denotados S y T con al menos dos alternativas. La probabilidad de que el individuo escoja una alternativa $a \in S$ cuando se enfrenta al conjunto de elección S se denotará $\mathbf{P}_S(a)$, suponiendo que para todo $S \subseteq C$, \mathbf{P}_S cumple con los axiomas de probabilidad sobre S . La probabilidad de que la alternativa

escogida en C pertenezca a S se denotará $\mathbf{P}_C(S)$, de forma tal que $\mathbf{P}_C(S) = \sum_{a \in S} \mathbf{P}_C(a)$. Por último, si $S = \{a, b\}$, entonces $\mathbf{P}(a, b)$ denotará a $\mathbf{P}_S(a)$.

El modelo de Luce

El modelo presentado por Luce lleva por nombre *axioma de elección*¹ y plantea una relación entre las probabilidades de elección definidas en los subconjuntos de C . El axioma de elección dice lo siguiente:

Axioma (R. D. Luce)

Para todo $S, T \subseteq C$ tales que $S \subseteq T$,

(i) si para $a \in S$ dado se tiene que para todo $b \in T$, $\mathbf{P}(a, b) \notin \{0, 1\}$, entonces

$$\mathbf{P}_T(a) = \mathbf{P}_T(S)\mathbf{P}_S(a);$$

(ii) si $\mathbf{P}(a, b) = 0$ para algún par $a, b \in T$, entonces para todo $S \subseteq T$ se tiene que

$$\mathbf{P}_T(S) = \mathbf{P}_{T \setminus \{a\}}(S \setminus \{a\}).$$

Estos simples axiomas imponen condiciones adicionales sobre las probabilidades de elección que no pueden ser derivadas sólo con la teoría de probabilidad. La parte (ii) dice que si alguna alternativa $a \in T$ nunca es escogida en una comparación de a pares, entonces a puede ser removida de T y esto no afecta a las probabilidades de elección de las otras alternativas en T . La parte (i) dice que la probabilidad de que a sea escogida en T no depende del subconjunto S de T que contenga a a en un proceso de dos etapas, en donde el individuo primero considera un subconjunto S de T y luego escoge un elemento particular $a \in S$.

Una vez realizado el proceso de eliminación antes explicado usando la parte (ii), Luce demuestra un resultado importante.

Teorema 1.1 (R. D. Luce, 1959)

Supóngase que $\mathbf{P}(a, b) \notin \{0, 1\}$ para todo $a, b \in C$. La parte (i) del axioma de elección se satisface si y sólo si existe una función $u : C \rightarrow (0, \infty)$ tal que

$$\mathbf{P}_S(a) = \frac{u(a)}{\sum_{b \in S} u(b)}. \quad (1.1)$$

Más aún, esta función es única salvo multiplicación por una constante positiva.

La función u puede ser interpretada como una «utilidad» determinista de la alternativa a . Como $u(a) > 0$, entonces $\mathbf{P}_S(a) > 0$; como $u(a) < \infty$, $\mathbf{P}_S(a) < 1$. Más aún, de (1.1) se ve que la probabilidad de que a sea escogida es creciente en su propia utilidad y decreciente en la utilidad de las otras alternativas, lo que es acorde a la intuición.

Se puede usar la transformación $v(a) = \ln(u(a))$ para reescribir (1.1) como

$$\mathbf{P}_S(a) = \frac{\exp(v(a))}{\sum_{b \in S} \exp(v(b))}, \quad (1.2)$$

¹*Choice axiom* en inglés; no confundir con el axioma de elección (*axiom of choice*) propuesto por Zermelo, herramienta básica y controversial para definir la teoría de conjuntos ZFC, base de la matemática actual.

que corresponde al famoso modelo logit multinomial explicado más adelante. Una implicación más fuerte del axioma de elección, que es menos atractiva, es una versión de la *independencia de alternativas irrelevantes* (IIA por sus siglas en inglés), que dice lo siguiente:

Proposición 1.2 (Independencia de Alternativas Irrelevantes)

Para todo $S, T \subseteq C$ tales que $S \subseteq T$ y para todo $a, b \in S$,

$$\frac{\mathbf{P}_S(a)}{\mathbf{P}_S(b)} = \frac{\mathbf{P}_T(a)}{\mathbf{P}_T(b)}. \quad (1.3)$$

Básicamente, (1.3) implica que la razón entre las probabilidades de elección de a y b no depende del conjunto que las contiene. Esto puede llevar a resultados contraintuitivos. Uno de los más conocidos es la «paradoja de los buses rojo y azul», adaptada del trabajo de Debreu [Deb60]: supóngase que un individuo debe llegar a destino y puede escoger entre un automóvil a o un bus b , de forma tal que $\mathbf{P}(a, b) = \frac{1}{2}$. Supóngase ahora que existen dos buses iguales salvo el color: uno es rojo (b_r) y otro azul (b_a). Sea $C = \{a, b_r, b_a\}$. Por último, supóngase que el color es irrelevante para el individuo, por lo que $\mathbf{P}_C(b_r) = \mathbf{P}_C(b_a)$. La intuición diría que $\mathbf{P}_C(a) = \frac{1}{2}$ y $\mathbf{P}_C(b_r) = \mathbf{P}_C(b_a) = \frac{1}{4}$, pero el axioma de elección permite demostrar que $\mathbf{P}_C(a) = \mathbf{P}_C(b_r) = \mathbf{P}_C(b_a) = \frac{1}{3}$. Esto es más chocante cuando se entiende que la probabilidad de escoger a se puede hacer arbitrariamente pequeña introduciendo una cantidad suficiente de buses de colores distintos.

Así, este modelo no considera casos en donde dos o más alternativas que para todos los efectos son iguales se muestran como alternativas distintas. El axioma de elección debe ser visto entonces como una primera aproximación a los procesos de decisión, con una agradable formulación de las probabilidades de elección (dada por el teorema 1.1).

El modelo de Tversky

Este modelo puede ser visto como un proceso estocástico en donde las alternativas son eliminadas sucesivamente hasta que sólo queda una. Cada alternativa es descrita por un conjunto de características (o aspectos) binarias (por ejemplo, «el computador tiene pantalla táctil»). Para lidiar con características que no sean intrínsecamente binarias, Tversky sugiere usar niveles umbrales (por ejemplo, «el computador tiene más/menos de 8 GB de RAM»).

El proceso de eliminación consiste en escoger una característica, descartar las alternativas que no posean dicha característica e iterar hasta que quede sólo una alternativa (o un conjunto de alternativas iguales, de donde se escoge de manera equiprobable). Esta es sólo la esencia del algoritmo; se verá en mayor profundidad en la sección 1.4. Nótese que es posible que muchas secuencias de eliminación llevan a escoger la misma alternativa, por lo que la probabilidad de elección debe reflejar esto.

Las probabilidades de elección quedan detalladas como sigue. Sea A el conjunto de las características. Se supone la existencia de una función $u : A \rightarrow (0, \infty)$ que especifica la «utilidad» de cada característica. Sea s el número de características después de haber eliminado las características comunes de las alternativas en el conjunto de elección $S \subseteq C$. Por último, sea $S_i \subseteq S$ el conjunto de alternativas que poseen la característica i , para $i \in \{1, \dots, s\}$. En el modelo de *eliminación por*

aspectos (EBA por sus siglas en inglés), la probabilidad de elegir la alternativa $a \in S$ está dada por

$$\mathbf{P}_S(a) = \sum_{i=1}^s \frac{u(i)}{\sum_{j=1}^s u(j)} \mathbf{P}_{S_i}(a). \quad (1.4)$$

Esta probabilidad es recursiva, por lo que es necesaria una condición terminal, que ya fue mencionada: cuando todas las características sean iguales en todas las alternativas de S , entonces se supone que $\mathbf{P}_S(a) = \frac{1}{|S|}$ para todo $a \in S$. Los pesos $\frac{u(i)}{\sum_{j=1}^s u(j)}$ representan la probabilidad de elegir la característica $i \in \{1, \dots, s\}$.

La probabilidad de elegir una alternativa está dada por una razón entre las «utilidades». En el modelo de Luce, específicamente en (1.1), esta razón se usa para calcular la probabilidad de escoger una alternativa. Esto sugiere que el modelo de Luce es un caso especial del modelo de Tversky. En particular, este modelo no impone la IIA, además de permitir alternativas con distintos grados de similaridad. Sin embargo, las características son esencialmente binarias, así que los valores absolutos de las características no tienen rol en la elección de una alternativa, y para características no binarias el individuo se enfrenta a umbrales minimales que le hacen rechazar o retener alternativas, lo que puede llevar a resultados contraintuitivos.

1.1.2. Modelos con utilidad estocástica

Contrastando a los modelos anteriores, los siguientes dos modelos son mucho más cercanos en espíritu a la teoría de elección neoclásica. Las dos principales interpretaciones del modelo con utilidad estocástica están adaptadas de los trabajos de psicología de Thurstone [Thu45] y de las formulaciones iniciadas por McFadden, adaptadas por economistas.

El modelo de Thurstone

El origen de la teoría de la utilidad aleatoria se remonta a los intentos de explicar resultados de experimentos psicológicos que involucraban comparaciones entre las intensidades de estímulos físicos. Para considerar la variabilidad en las respuestas, Thurstone propone un modelo basado en dos ideas: primero, un estímulo dado provoca una «sensación» que es *la realización de una variable aleatoria*, y segundo, las respuestas de un individuo que compara dos estímulos resulta en *la comparación de la realización de dos variables aleatorias* que representan las sensaciones provocadas por el estímulo. Para ligar esto con la teoría económica, se puede pensar que las utilidades varían en distintos momentos, y el proceso de decisión consiste en la idea de escoger la alternativa con la utilidad más grande momentánea.

Bajo el enfoque anterior, las valoraciones de las alternativas en C se consideran como variables aleatorias $u_1 + \varepsilon_1, \dots, u_n + \varepsilon_n$ (con $n = |C|$), que son comparadas por el individuo a la hora de su decisión. Las variables u_1, \dots, u_n son valores de escala constantes asociados a las alternativas, y $\varepsilon_1, \dots, \varepsilon_n$ son variables aleatorias. Se supone que la función de distribución acumulativa F de $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ es tal que $F \ll \lambda$, con λ la medida de Lebesgue (para que $\mathbb{P}(\varepsilon_i - \varepsilon_j = \alpha) = 0$ para cualquier α e $i \neq j$). Si los ε_i son tales que $\mathbb{E}(\varepsilon_i) = 0$ (de no cumplirse, la media se puede sumar a u_i), entonces la probabilidad de elección se puede construir de la siguiente manera:

$$\mathbf{P}_C(i) = \mathbb{P} \left[u_i + \varepsilon_i = \max_{j=1, \dots, n} (u_j + \varepsilon_j) \right], \quad i \in \{1, \dots, n\}. \quad (1.5)$$

La interpretación económica

Este modelo es un poco distinto a los otros tres anteriores, pues se considera ahora una población de individuos enfrentados al mismo conjunto de elección C , y el objetivo es determinar la fracción de la población que escoge una alternativa dada. La población es dividida en subpoblaciones, con cada subpoblación homogénea con respecto a ciertos factores socioeconómicos observables. Cada individuo posee una función de utilidad determinista $U : C \rightarrow \mathbb{R}$.

La aleatoriedad proviene del modelador, que sólo puede observar imperfectamente las características que influyen la elección del individuo y, por ende, no conoce U en su totalidad. Así, la función U se descompone en dos partes: una función u que representa la parte conocida de la utilidad y definida sobre características observables, y una función e , la diferencia entre U y u . Para $i \in \{1, \dots, n\}$, la utilidad de la alternativa i se escribe como $U_i = u_i + e_i$.

Si bien el comportamiento del individuo es determinista en este enfoque, el modelador no es capaz de predecir dicho comportamiento pues no puede observar e . Se supone que cada miembro difiere de los demás dentro de su subpoblación sólo en factores no observables, por lo que el modelador puede a lo más predecir la elección del individuo salvo una función de probabilidad. Por consiguiente, se representa a la valoración e_i con una variable aleatoria ε_i con media cero, y la utilidad U_i se modela por la variable aleatoria $\tilde{U}_i = u_i + \varepsilon_i$.

Aquí u_i es la llamada *utilidad observable* que refleja las preferencias de la subpoblación por la i -ésima alternativa, y ε_i toma en consideración las diferencias de gusto de los miembros de dicha subpoblación. Por último, la probabilidad de que un individuo escogido al azar escoja la alternativa i se construye del principio de maximización de utilidades individuales, dada por

$$\mathbf{P}_C(i) = \mathbb{P} \left[\tilde{U}_i = \max_{j=1, \dots, n} \tilde{U}_j \right], \quad i \in \{1, \dots, n\}. \quad (1.6)$$

El enfoque estocástico de este modelo es distinto al de los demás. Aquí la regla de decisión y la utilidad del individuo son ambas deterministas. La aleatoriedad viene de la falta de información disponible para el modelador. Según Manski [Man77], hay cuatro distintas fuentes de incertidumbre: características no observables, variaciones no observables en las utilidades individuales, errores de medición y mala especificación funcional.

1.1.3. Identificación de los Modelos

Un aspecto importante en los modelos de elección es la identificación de los mismos, que puede resumirse en las siguientes dos frases: «sólo importan las diferencias en las utilidades» y «la escala total de la utilidad es arbitraria». Sólo por esta subsección se sigue el libro de Train [Tra09, capítulo 2].

La primera frase se refiere a que el valor absoluto de la utilidad no es relevante al comportamiento del individuo ni al modelo del investigador. Si se añade una constante a la utilidad de todas las alternativas, la alternativa con la utilidad más grande no cambia. Esto implica en general que los únicos parámetros que se pueden estimar son los que capturan diferencias entre alternativas. La solución a esta situación es fijar una utilidad (cualquiera) u_i en un valor fijo arbitrario.

Por otro lado, la segunda frase se refiere a que tampoco es relevante multiplicar las utilidades por una constante positiva. La alternativa con la utilidad más grande sigue siendo la elegida. Para

dar cuenta de este hecho se normaliza la escala de la utilidad, lo que, por lo general, se realiza normalizando la varianza de los términos de error.

1.2. Las probabilidades de elección en los modelos de elección discretos multinomiales

Anderson *et al.* dan una demostración al caso general cuando hay $n \geq 2$ alternativas, n variables aleatorias $\varepsilon_1, \dots, \varepsilon_n$ con distribución acumulativa $F(x_1, \dots, x_n)$ definida en \mathbb{R}^n , con la utilidad de la alternativa i dada por $\tilde{U}_i = u_i + \varepsilon_i$ y $\mathbb{E}(\varepsilon_i) = 0$. Se supone que F posee una densidad f tal que $\text{sop}(f) \subset \mathbb{R}^n$ es convexo.

De (1.6) (o de (1.5)) se puede reescribir la probabilidad de escoger i como

$$\mathbf{P}_C(i) = \mathbb{P}(\varepsilon_1 - \varepsilon_i \leq u_i - u_1, \dots, \varepsilon_n - \varepsilon_i \leq u_i - u_n), \quad i \in \{1, \dots, n\}.$$

Si los ε_i son i.i.d., entonces la expresión anterior se puede manipular fácilmente para llegar a

$$\mathbf{P}_C(i) = \int_{-\infty}^{\infty} f(x) \prod_{j \neq i} F(u_i - u_j + x) dx, \quad i \in \{1, \dots, n\}. \quad (1.7)$$

La idea de (1.7) es la siguiente: para una realización x de ε_i , la probabilidad de elegir i es el producto dentro de la integral anterior por ser i.i.d. Integrando todas las posibilidades de x se llega a (1.7). Cuando los ε_i no son i.i.d. se debe trabajar un poco más para llegar a lo siguiente:

$$\mathbf{P}_C(i) = \int_{-\infty}^{\infty} \frac{\partial F}{\partial x_i}(u_i - u_1 + x, \dots, x, \dots, u_i - u_n + x) dx, \quad i \in \{1, \dots, n\}. \quad (1.8)$$

Intuitivamente (1.8) se explica así: para cada x el integrando es la densidad de probabilidad de que ε_i tome el valor de x y de que $\varepsilon_j \leq u_i - u_j + x$, para $j \neq i$. Integrando todos los valores de x se llega a la probabilidad de elección.

1.3. El logit multinomial

La probabilidad en (1.2) es bastante usada y es parte del presente trabajo. El siguiente resultado, atribuido por Luce y Suppes [LS65] a Holman y Marley en un artículo no publicado, especifica la distribución de los ε_i que generan (1.2).

Teorema 1.3 (Holman y Marley)

Supóngase que los ε_i son i.i.d. con distribución Gumbel (o doble exponencial) de media 0,

$$F(x) = \mathbb{P}(\varepsilon_i \leq x) = \exp \left[-\exp \left(-\frac{x}{\mu} - \gamma \right) \right], \quad (1.9)$$

donde γ es la constante de Euler-Mascheroni ($\gamma \approx 0,5772$) y $\mu > 0$ es constante. Entonces las probabilidades de elección resultantes están dadas por

$$\mathbf{P}_C(i) = \frac{\exp(u_i/\mu)}{\sum_{j=1}^n \exp(u_j/\mu)}, \quad i \in \{1, \dots, n\}.$$

Yellot da una caracterización más fuerte entre la doble exponencial y el logit multinomial.

Teorema 1.4 (Yellot, 1977)

Supóngase que $n \geq 3$, que los ε_i son i.i.d., y que la función de distribución común es estrictamente creciente en \mathbb{R} . Entonces las probabilidades de elección están dadas por el modelo logit multinomial si y sólo si los ε_i están distribuidos como una doble exponencial.

Este resultado es interesante de por sí, puesto que la probabilidad de elección toma una forma bastante sencilla y útil. Existen otras caracterizaciones del modelo igual de interesantes, así como generalizaciones del modelo logit. El *logit anidado*, por ejemplo, modela la elección en dos etapas, mientras que el *valor extremo generalizado* considera una familia más grande de distribuciones (que contiene a la distribución Gumbel) para generar las probabilidades de elección. Las propiedades de la distribución Gumbel se pueden ver en el apéndice A.

1.4. El modelo EBA como modelo logit

El contexto de esta tesis es el modelo EBA propuesto por Tversky [Tve72a, Tve72b]. Si bien este modelo ya fue explicado en la subsección 1.1.1 siguiendo el libro de Anderson *et al.* [AdT92], la base de la revisión del modelo hecha a continuación es el artículo de Kohli *et al.* [KJM14] en donde, por un lado, se formaliza el modelo con un algoritmo y se desarrolla una teoría de error para el modelo en la cual la distribución doble exponencial juega un rol crucial.

Sea C un conjunto de elección y sea A el conjunto de aspectos usados para describir a las alternativas. Sea $n = |A|$ el número de aspectos y sea $\alpha_k > 0$ un número que denote el peso de importancia que un individuo asocia al aspecto $k \in A$. El perfil de una alternativa $i \in C$ se representa por un vector $\mathbf{x}_i = (x_{i1}, \dots, x_{in})$, en donde x_{ik} es una variable binaria que vale 1 si y sólo si la alternativa i posee el aspecto k . Se usa indistintamente que $i \in C$ o que $\mathbf{x}_i \in C$ para simplificar notación. El proceso de eliminación puede ser descrito por el siguiente algoritmo.

Algoritmo 1 Eliminación por Aspectos.

Entrada: Conjunto de elección C , aspectos A , pesos $(\alpha_k)_{k=1}^n$, perfiles $(\mathbf{x}_i)_{i \in C}$.

Salida: Alternativa $i \in C$.

- 1: $A_0 \leftarrow A, C_0 \leftarrow C, j \leftarrow 1$.
 - 2: **repetir**
 - 3: Fijar el conjunto A_j de aspectos que están en al menos una alternativa, pero no en todas:
 $A_j \leftarrow \{k \in A_{j-1} : 0 < \sum_{i \in C_{j-1}} x_{ik} < |C_{j-1}|\}$.
 - 4: Escoger un aspecto $k_j \in A_j$ con probabilidad $q(k_j, A_j) = \frac{\alpha_{k_j}}{\sum_{k \in A_j} \alpha_k}$.
 - 5: Eliminar alternativas sin aspecto k_j : $C_j \leftarrow \{i \in C_{j-1} : x_{ik_j} = 1\}$.
 - 6: **si** $|C_j| > 1$ **entonces**
 - 7: $j \leftarrow j + 1$.
 - 8: **fin si**
 - 9: **hasta que** $|C_j| = 1$
 - 10: **devolver** C_j .
-

Una *instancia* de EBA es una secuencia particular de aspectos usados para realizar un proceso de eliminación. Toda instancia definida sobre n aspectos termina en $m \leq n$ pasos. Con todo esto,

la probabilidad de usar una instancia particular (k_1, \dots, k_m) para eliminar está dada por

$$\mathbb{P}(\text{usar } (k_1, \dots, k_m)) = \prod_{j=1}^m q(k_j, A_j) = \prod_{j=1}^m \frac{\alpha_{k_j}}{\sum_{k \in A_j} \alpha_k}. \quad (1.10)$$

Cabe destacar que distintas instancias pueden llevar a elegir la misma alternativa.

Kohli *et al.* desarrollan una teoría de error para el modelo EBA. Considerando los modelos con utilidad estocástica de la subsección 1.1.2, un individuo le asocia una utilidad aleatoria a cada aspecto $k \in A$ de la forma $U_k = u_k + \varepsilon_k$, con u_k una componente determinista y ε_k un error asociado. Si bien el modelo EBA permite que distintas instancias puedan ser usadas en distintas ocasiones de elección, se supone que las utilidades de los aspectos sólo pueden cambiar entre ocasiones de elección.

Sin perder generalidad, supóngase que $U_1 > \dots > U_n$ en una ocasión dada. Sea $s = (1, \dots, n)$ la secuencia de aspectos ordenados de manera decreciente en utilidades. Un individuo usa s de forma lexicográfica como sigue: primero elimina las alternativas sin el aspecto 1, de las restantes elimina las que no poseen el aspecto 2, y así sucesivamente hasta que quede una alternativa. Esta regla puede o no usar todos los aspectos para eliminar, ya sea porque todas o ninguna de las alternativas poseen un aspecto o porque el proceso puede parar antes de usar todos los aspectos.

El objetivo del artículo de Kohli *et al.* es especificar una distribución estadística para cada ε_k que posea las siguientes propiedades:

1. dada la secuencia $s = (1, \dots, n)$, supóngase que un individuo usa los aspectos k_1, \dots, k_m en ese orden para efectivamente eliminar alternativas, con $1 \leq m \leq n$. Si $m < n$, entonces existen otras secuencias que también eliminan alternativas usando k_1, \dots, k_m en ese orden. Sea S el conjunto de tales secuencias y $\mathbf{P}(s)$ la probabilidad de usar la secuencia $s \in S$. La teoría de error debe asegurar que la suma de las probabilidades $\mathbf{P}(s)$ sobre S sea igual a la probabilidad dada por Tversky:

$$\mathbf{P}(S) = \sum_{s \in S} \mathbf{P}(s) = \prod_{j=1}^m \frac{\alpha_{k_j}}{\alpha_{k_j} + \dots + \alpha_n};$$

2. debe ser posible particionar el conjunto de las $n!$ secuencias en conjuntos S_1, \dots, S_r tales que $\mathbf{P}(S_1), \dots, \mathbf{P}(S_r)$ especifiquen exhaustivamente las probabilidades de todas las instancias de EBA posibles.

Se demuestra en dicho artículo que si ε_k sigue una distribución Gumbel estándar entonces se cumplen estas propiedades con las probabilidades de elección dadas por

$$\mathbf{P}(S) = \prod_{j=1}^m \frac{\exp(u_{k_j})}{\exp(u_{k_j}) + \dots + \exp(u_n)},$$

y considerando un modelo EBA con $\alpha_k = \exp(u_k)$, para todo $k \in A$, entonces se recuperan las probabilidades dadas por (1.10). Por último, se da una expresión cerrada para las probabilidades de elección, de manera intuitiva cuando hay dos, tres y cuatro alternativas, y su extensión general a n alternativas.

1.5. Diseño de experimentos: cuestionarios eficientes

En muchos campos, como el *marketing*, la planificación de transportes, la política y pronóstico de consumo de energía y modelos de evacuación, uno de los objetivos en el ámbito de modelar las elecciones discretas es poder rescatar las valoraciones deterministas de las utilidades (el término u de la expresión $U = u + \varepsilon$) para realizar estudios de interés, como maximización de la utilidad u optimización y estadística aplicada a la teoría del consumidor. Es en este contexto en donde surgen distintos métodos para poder obtener datos, reales o simulados, que puedan ser usados para poder estimar las valoraciones subyacentes.

Uno de los métodos más usados es el diseño de experimentos. En su forma más simple, un experimento apunta a predecir el resultado al introducir un cambio en las condiciones prefijadas. Más en concreto, el concepto de «diseño de experimentos» se refiere a definir objetos a observar, tratamientos a aplicar, y una función que asigne dichos tratamientos a dichos objetos [Bai08], [DMSB15, capítulo 10].

En el contexto de la teoría de elección discreta, el diseño de experimento lleva el nombre de *experimento de elección*, en donde los objetos a observar son los atributos o características que definen a las alternativas y los tratamientos son las mismas alternativas y los conjuntos de elección. El diseño consiste en confeccionar los conjuntos de elección, compuestos de varias alternativas, y la información a medir es la elección de una alternativa [Kuh10]. Indistintamente se usan las palabras «diseño» y «cuestionario» para referirse al mismo concepto.

El siguiente objetivo del diseño de experimentos es poder medir qué tan eficiente es un cuestionario a la hora de estimar parámetros. Para los modelos lineales de la forma $y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, donde y es un vector de observaciones, \mathbf{X} una matriz de (funciones de) variables explicativas, $\boldsymbol{\beta}$ un vector de parámetros desconocidos y $\boldsymbol{\varepsilon}$ un vector de errores, abundan los criterios de eficiencia, y el modelo logit cae dentro de estos modelos. Dado el contexto de utilidad estocástica, McFadden [McF73] supone que la utilidad del individuo n que escoge la alternativa i está dado por

$$U_{ni} = \mathbf{z}_{ni}^T \boldsymbol{\beta} + \varepsilon_{ni}, \quad (1.11)$$

donde $\mathbf{z}_{ni} = z(\mathbf{x}_{ni}, \mathbf{s}_n)$ es un vector de variables observables que depende de los aspectos de la alternativa (\mathbf{x}_{ni}) y de la persona (\mathbf{s}_n), $\boldsymbol{\beta}$ es un vector de coeficientes a estimar y ε_{ni} es el error asociado. Nótese que en (1.11) el término determinista $\mathbf{z}_{ni}^T \boldsymbol{\beta}$ juega el rol de u_i visto en la subsección 1.1.2 (salvo índices que denotan individuos). Este modelo es el llamado *compensatorio*, pues el vector $\boldsymbol{\beta}$ hace que todas las variables observables \mathbf{z}_{ni} sean consideradas «bajo una misma moneda». Así, atributos «malos» pueden ser compensados con atributos «buenos».

Todos los criterios se basan en la siguiente idea: se sabe que el estimador de mínimos cuadrados minimiza la varianza (bajo las condiciones del teorema de Gauss-Markov). En una variable, el recíproco de la varianza de un estimador («eficiente») se llama la información de Fisher. Dada esta reciprocidad, «minimizar la varianza» corresponde a «maximizar la información». En más variables, la información de Fisher es una matriz, por lo que «maximizar la información» se traduce en maximizar una función a valores reales de dicha matriz.

A continuación se destacan los criterios más usados. Bunch, Louviere y Anderson [BLA94] y Kuhfeld, Tobias y Garratt [KTG94] proponen, dada la matriz de diseño \mathbf{X} (en donde las filas corresponden a los vectores \mathbf{z}_{ni}), optimizar una función escalar normalizada de la matriz de información $\mathbf{X}^T \mathbf{X}$, como la traza de la inversa (A -eficiencia), el determinante (D -eficiencia) o

el error estándar más grande de la predicción (G -eficiencia). Huber y Zwerina [HZ96] muestran cuatro propiedades heurísticas de diseños eficientes, que no cambian mucho respecto a las medidas antes descritas: balance de nivel, ortogonalidad, nivel de traslape mínimo y balance de utilidad. Sándor y Wedel [SW01] extienden el trabajo anterior para un marco bayesiano. Toubia y Hauser [TH07] complementan estas técnicas generalizando las medidas de la matriz de información y definiendo la M -eficiencia.

Además de los trabajos antes citados, se han estudiado generalizaciones del modelo logit para dar cuenta de más características del mismo. Por citar algunos, Sándor y Wedel [SW02] construyen diseños óptimos cuando $\beta \sim \mathcal{N}(\mu, \Sigma)$, reflejando heterogeneidad en los individuos; Liu y Arora [LA11] generalizan un modelo en dos pasos, donde primero se eliminan ciertas alternativas y luego se escoge, dentro de las alternativas restantes, siguiendo el modelo logit; Crabbe, Akinç y Vandebroek [CAV13] obtienen diseños eficientes individuales que consideran las respuestas de las preguntas anteriores. Sagnol y Harman [SH15] muestran que, en modelos de regresión lineal, el criterio de la D -optimalidad referente a la «calidad de la estimación de los parámetros» del diseño es representable por conos de segundo orden y, por ende, el método de programación cónica de segundo orden puede ser usado para calcular un diseño D -óptimo aproximado; y Duarte y Wong [DW15] usan programación semidefinida positiva para construir diseños óptimos bayesianos para modelos de regresión no lineales, extendiendo la formulación ya existente para modelos lineales.

La construcción de diseños eficientes para un modelo de elección discreta es un problema frecuente en *marketing* e investigación de operaciones. En particular, el marco de esta tesis es el desarrollo de cuestionarios eficientes suponiendo que la elección discreta subyacente se rige bajo el modelo EBA. Dicho modelo no ha sido tan estudiado como los modelos compensatorios, dada su complejidad y difícil manipulación, aún cuando existe evidencia de que EBA generaliza al modelo logit multinomial [KJM14]. Este problema no ha sido resuelto aún y sirve de motivación para el presente trabajo.

Capítulo 2

Un algoritmo de optimización para el desarrollo de cuestionarios con dos alternativas

Durante este capítulo y el siguiente se considera el caso en donde las probabilidades de elección son conocidas. Dado un conjunto de elección cualquiera C y un conjunto de aspectos A de tamaño n con el cual se construyen las alternativas de C , entonces para toda alternativa $i \in C$ la probabilidad $\mathbf{P}_C(\mathbf{x}_i)$ de elegir la alternativa i es conocida y tiene la forma dada por Kohli *et al.* [KJM14].

El problema que se busca resolver es el siguiente: dado un conjunto de aspectos A y un número entero fijo $T \geq 1$, se desea diseñar un cuestionario que consista de a lo más T preguntas con el fin de recuperar la mayor cantidad de valoraciones u_k de los aspectos $k \in A$. Posteriormente se miden métricas estadísticas para discernir qué tan potentes son dichos cuestionarios.

Para resolver este problema se utilizan herramientas de la programación lineal mixta (MIP por sus siglas en inglés). En concreto, se modela el problema como un problema MIP binivel en donde se busca optimizar una función que refleje si el modelo es identificable o no. Se entiende por «identificable» la existencia de un único vector de valoraciones compatible con el número de preguntas T y las probabilidades de elección dadas. Surge la naturaleza de los problemas de tipo MIP pues el diseño se codifica con variables discretas, mientras que las valoraciones son variables continuas.

El comienzo del trabajo se da en su caso más sencillo: sólo hay dos alternativas por conjunto de elección. Un primer paso es fijar notación con respecto a los aspectos que forman las alternativas, para poder ser consistentes con lo planteado por Tversky y con el trabajo realizado por Kohli *et al.*

Las alternativas de un conjunto de elección están definidas sobre un conjunto de atributos en común. Cada alternativa debe poseer todos los atributos. Cada atributo posee niveles, y dos (o más) alternativas se diferencian entre sí por los niveles que ellas poseen de los distintos atributos. Se denota por G a un atributo y por \mathcal{G} al conjunto de todos los atributos. Se denota por $k \in G$ al nivel k -ésimo del atributo G . Cada atributo debe tener al menos dos niveles, pues si sólo posee uno entonces todas las alternativas poseen el atributo en el mismo nivel, y por ende es descartado en este modelo.

Ejemplo (Cámaras fotográficas)

Supóngase que se desea representar cámaras fotográficas con ciertos atributos. Cada cámara se describe con cinco atributos (precio, resolución, vida de la batería, enfoque óptico y tamaño). Los distintos niveles son los presentados a continuación:

- precio: \$500, \$400, \$300 y \$200;
- resolución: 2, 3, 4 y 5 megapíxeles;
- vida de la batería: 150, 300, 450 y 600 fotografías;
- enfoque óptico: 2×, 3×, 4× y 5×, y
- tamaño: SLR, media, de bolsillo y ultracompacta.

Sea $P = \{\$500, \$400, \$300, \$200\}$ el conjunto de todos los aspectos de precio. Entonces $P \in \mathcal{G}$. De manera análoga, si R, V, E, T son los conjuntos de aspectos de resolución, vida, enfoque y tamaño, entonces $R, V, E, T \in \mathcal{G}$. El primer nivel de P es \$500, el segundo nivel es \$400 y así sucesivamente.

En la notación de Kohli *et al.*, $A = \bigcup_{G \in \mathcal{G}} G$. Indistintamente, se llamará «grupo» o «atributo» a $G \in \mathcal{G}$, y «aspecto» o «nivel de atributo» a $k \in G$, para todo $G \in \mathcal{G}$.

Otra notación bastante usada en el campo para explicar el contexto de atributos y niveles en el cual se trabaja es el producto $\prod_{G \in \mathcal{G}} n_G$, donde n_G es el número de niveles del atributo G . Esto permite abstraer el tipo de atributo: no es necesario conocer qué describe el atributo, sólo sus niveles. Como el orden no influye para describir al objeto a modelar, por convención, la notación se usa en orden ascendente de niveles, sin importar los empates.

Ejemplo revisado (Cámaras fotográficas)

Las cámaras fotográficas siguen la notación $4 \times 4 \times 4 \times 4 \times 4$, o 4^5 . Si los precios sólo fueran $\{\$500, \$400\}$, la vida $\{300, 450\}$ y el enfoque $\{3\times, 4\times, 5\times\}$, la notación cambia a $2^2 \times 3 \times 4^2$.

2.1. Formato de las preguntas

Siguiendo y adaptando la notación de Kohli *et al.*, sean \mathbf{x}_{1t} y \mathbf{x}_{2t} los perfiles de las dos alternativas del conjunto de elección mostrado en la pregunta t , para $t \in [T] := \{1, \dots, T\}$ y T un número fijo de preguntas. Cada componente binaria de \mathbf{x}_{it} se indexa como sigue:

$$x_{itkG} = \begin{cases} 1 & \text{si el aspecto } k \in G (G \in \mathcal{G}) \text{ está en la alternativa } i \text{ en la pregunta } t, \\ 0 & \text{en otro caso.} \end{cases} \quad (2.1)$$

Los perfiles \mathbf{x}_{1t} y \mathbf{x}_{2t} deben ser capaces de modelar alternativas que se condigan con la realidad. Para todo atributo G , no puede ocurrir que existan dos variables $x_{itkG} = 1 = x_{itjG}$ para $k, j \in G$ y $k \neq j$; de otra forma la alternativa estaría diciendo que un atributo posee dos niveles a la vez. Dado esto, cada perfil debe cumplir la igualdad

$$\sum_{k \in G} x_{itkG} = 1, \quad i \in \{1, 2\}, G \in \mathcal{G}, t \in [T]. \quad (2.2)$$

Sea $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})$ para $i \in \{1, 2\}$. Se denota por $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ al diseño, que consta de las T preguntas y de las dos alternativas mostradas en cada pregunta.

2.2. Las probabilidades

Se analizan ahora las probabilidades de elección. Se debe recordar que los aspectos que ambas alternativas tienen en común no son relevantes a la hora de escoger. Sea entonces $A_{it} \subseteq A$ el conjunto de aspectos que aparecen únicamente en la alternativa i en la pregunta t , con $i \in \{1, 2\}$ y $t \in [T]$. Denotando $z_{kG} = \exp(u_{kG})$ (y entendiendo que k y G se corresponden), se tiene en este caso que las probabilidades de elección dadas por Kohli *et al.* tienen la forma

$$p(\mathbf{x}_{it}) = \mathbf{P}_{\{1,2\}}(\text{elegir alternativa } i \text{ en la pregunta } t) = \frac{\sum_{G \in \mathcal{G}} \sum_{k \in A_{it}} z_{kG}}{\sum_{G \in \mathcal{G}} \sum_{k \in A_{1t} \cup A_{2t}} z_{kG}}. \quad (2.3)$$

De esto se desprende un sistema lineal para las variables z_k . En concreto,

$$\begin{aligned} (1 - p(\mathbf{x}_{1t})) \sum_{G \in \mathcal{G}} \sum_{k \in A_{1t}} z_{kG} - p(\mathbf{x}_{1t}) \sum_{G \in \mathcal{G}} \sum_{k \in A_{2t}} z_{kG} &= 0, \quad t \in [T], \\ (1 - p(\mathbf{x}_{2t})) \sum_{G \in \mathcal{G}} \sum_{k \in A_{2t}} z_{kG} - p(\mathbf{x}_{2t}) \sum_{G \in \mathcal{G}} \sum_{k \in A_{1t}} z_{kG} &= 0, \quad t \in [T]. \end{aligned} \quad (2.4)$$

Ambas ecuaciones son en realidad una sola, pues $1 - p(\mathbf{x}_{1t}) = p(\mathbf{x}_{2t})$. Como el diseño consta de T preguntas, entonces por cada pregunta se tiene una igualdad del estilo (2.4). Esto sugiere usar una estructura matricial de la forma $\mathbf{A}\mathbf{z} = \mathbf{b}$, con \mathbf{A} y \mathbf{b} de dimensiones adecuadas que reflejen el sistema lineal.

Dado que las probabilidades de elección sólo dependen de los aspectos únicos de cada alternativa, es necesario construir a partir del diseño $(\mathbf{x}_1, \mathbf{x}_2)$ ciertos vectores $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$ que reflejen esta «unicidad» de aspectos, es decir,

$$y_{itkG} = \begin{cases} 1 & \text{si el aspecto } k \in G \ (G \in \mathcal{G}) \text{ está únicamente en la alternativa } i \text{ en la pregunta } t, \\ 0 & \text{en otro caso.} \end{cases}$$

La construcción de estos vectores $(\mathbf{y}_1, \mathbf{y}_2)$ es bastante fácil. Las relaciones lógicas que permiten esta construcción y su equivalencia a restricciones lineales se muestran a continuación, para todo $k \in G$, $G \in \mathcal{G}$, $t \in [T]$.

$$\begin{aligned} (x_{1tkG} = 0 \implies y_{1tkG} = 0) &\iff y_{1tkG} \leq x_{1tkG}, \\ (x_{1tkG} = 1 \wedge x_{2tkG} = 0 \implies y_{1tkG} = 1) &\iff y_{1tkG} \geq x_{1tkG} - x_{2tkG}, \\ (x_{2tkG} = 0 \implies y_{2tkG} = 0) &\iff y_{2tkG} \leq x_{2tkG}, \\ (x_{2tkG} = 1 \wedge x_{1tkG} = 0 \implies y_{2tkG} = 1) &\iff y_{2tkG} \geq x_{2tkG} - x_{1tkG}, \\ (x_{1tkG} = 1 = x_{2tkG} \implies y_{1tkG} = 0 = y_{2tkG}) &\iff y_{1tkG} + y_{2tkG} \leq 4 - 2x_{1tkG} - 2x_{2tkG}. \end{aligned}$$

Para simplificar, estas relaciones se denotan $R(\mathbf{x}, \mathbf{y})$. Nótese que, si bien se declaran las variables y_{itkG} como binarias, el hecho de que las variables x_{itkG} lo sean fuerza a que todo y_{itkG} tome un valor en $\{0, 1\}$. Por ende, para simplificar la posterior resolución, se declara que y_{itkG} es continua en $[0, 1]$.

2.3. El problema MIP binivel

Un detalle a resolver es el siguiente: dado el número de preguntas T , ¿cuántos vectores \mathbf{z} son compatibles con el modelo? Nada hace señalar de buenas a primeras que siempre habrá un

único vector acorde a las probabilidades y al número de preguntas. Puede ocurrir que existan dos vectores \mathbf{z}, \mathbf{z}' que cumplan con las restricciones del tipo (2.4) para el mismo diseño y tales que $\|\mathbf{z} - \mathbf{z}'\|$ sea muy grande (para cualquier norma), es decir, que las valoraciones sean distintas.

Para intentar resolver este detalle se usa la estructura de los problemas MIP binivel [BM73, CN77]. A grandes rasgos, los MIP binivel son una clase de problemas de optimización mixta (llamados de nivel superior) en donde una de las restricciones es maximizar o minimizar la función objetivo de otro problema de optimización (llamado de nivel inferior).

La intuición detrás de usar este tipo de problemas para el presente trabajo es la siguiente:

1. dado un diseño de T preguntas, el problema de nivel inferior busca maximizar la distancia entre dos vectores \mathbf{z} y \mathbf{z}' sujetos a la restricción (2.4) (suponiendo conocidas las probabilidades de elección). El objetivo es forzar a que los dos vectores sean distintos;
2. el problema de nivel superior busca minimizar el valor máximo encontrado en el nivel inferior sujeto a que el diseño de las preguntas sea uno factible. Si dicho mínimo es 0, entonces sólo hay un vector \mathbf{z} que recupera las valoraciones. Se retorna ese vector y el diseño asociado.

Se muestra a continuación el problema de nivel inferior. Como se menciona antes, las variables del problema serán dos vectores \mathbf{z}, \mathbf{z}' , ambos compatibles con (2.4). La función objetivo del problema refleja la distancia entre ambos vectores, a saber,

$$\|\mathbf{z} - \mathbf{z}'\|_1 = \sum_{G \in \mathcal{G}} \sum_{k \in G} |z_{kG} - z'_{kG}|. \quad (2.5)$$

Si bien ninguna norma es lineal, el módulo es linealizable fácilmente (ver sección 2.4.2). Por eso se escoge la norma-1 por sobre las otras.

Una de las restricciones que han de satisfacer estos vectores es que cumplan con la igualdad (2.4). Usando la notación introducida anteriormente, las restricciones quedan

$$\sum_{G \in \mathcal{G}} \sum_{k \in G} [p(\mathbf{x}_{2t})z_{kG}y_{1tkG} - p(\mathbf{x}_{1t})z_{kG}y_{2tkG}] = 0, \quad t \in [T], \quad (2.6)$$

$$\sum_{G \in \mathcal{G}} \sum_{k \in G} [p(\mathbf{x}_{2t})(z_{kG} - z'_{kG})y_{1tkG} - p(\mathbf{x}_{1t})(z_{kG} - z'_{kG})y_{2tkG}] = 0, \quad t \in [T]. \quad (2.7)$$

La restricción dice que un vector (sin perder generalidad, \mathbf{z}) cumpla la igualdad (2.4) y que en conjunto ambos vectores lo hagan. Nótese que se reemplaza $1 - p(\mathbf{x}_{1t})$ por $p(\mathbf{x}_{2t})$.

Otra restricción a considerar tiene que ver con la identificación del problema, que permite fijar una valoración arbitraria a un valor arbitrario. Lo más sencillo es fijar alguna (cualquiera) en 0. Esto se traduce en

$$z_{1H} = 1 \quad \text{y} \quad z'_{1H} = 1, \quad (2.8)$$

donde H es el primer atributo, escogido arbitrariamente.

La última restricción ha de reflejar la naturaleza de las variables z_{kG} . Como éstas representan a exponenciales de valoraciones (que a su vez son números reales) la restricción natural a considerar sería $\mathbf{z}, \mathbf{z}' > \mathbf{0}$. Sin embargo, dichas restricciones no son *ad hoc* a los problemas de optimización, pues la desigualdad es estricta. Se considera entonces la siguiente simplificación:

$$\mathbf{0} \leq \mathbf{z}, \mathbf{z}' \leq M\mathbf{1}, \quad (2.9)$$

en donde M es una constante grande adecuada y $\mathbf{1}$ es el vector con todas sus componentes iguales a 1. Usando (2.5) como función objetivo y (2.6) – (2.9) como restricciones, el problema de nivel inferior queda explicitado como sigue:

$$\begin{aligned}
f(\mathbf{x}, \mathbf{y}) &= \max \sum_{G \in \mathcal{G}} \sum_{k \in G} |z_{kG} - z'_{kG}| \\
\text{sujeto a } &\sum_{G \in \mathcal{G}} \sum_{k \in G} [p(\mathbf{x}_{2t})z_{kG}y_{1tkG} - p(\mathbf{x}_{1t})z_{kG}y_{2tkG}] = 0, & t \in [T], \\
&\sum_{G \in \mathcal{G}} \sum_{k \in G} [p(\mathbf{x}_{2t})(z_{kG} - z'_{kG})y_{1tkG} - p(\mathbf{x}_{1t})(z_{kG} - z'_{kG})y_{2tkG}] = 0, & t \in [T], \\
&z_{1H} = 1, \\
&z'_{1H} = 1, \\
&\mathbf{0} \leq \mathbf{z}, \mathbf{z}' \leq M\mathbf{1}.
\end{aligned} \tag{P_{x,y}}$$

Como se menciona más arriba, este problema supone la existencia de un diseño $(\mathbf{x}_1, \mathbf{x}_2)$ que es usado como dato (reflejado en las probabilidades de elección) para poder calcular dos vectores de valoraciones.

Ahora se plantea el problema de nivel superior. En este problema las variables de decisión son las que conforman el diseño, y se busca minimizar el valor máximo encontrado en $(P_{x,y})$. Juntando las restricciones (2.1) y (2.2), el problema de nivel superior queda como sigue:

$$\begin{aligned}
&\min_{\mathbf{x}, \mathbf{y}} f(\mathbf{x}, \mathbf{y}) \\
\text{sujeto a } &\sum_{k \in G} x_{itkG} = 1, & i \in \{1, 2\}, G \in \mathcal{G}, t \in [T], \\
&R(\mathbf{x}, \mathbf{y}), \\
&x_{itkG} \in \{0, 1\}, & i \in \{1, 2\}, k \in G, G \in \mathcal{G}, t \in [T], \\
&0 \leq y_{itkG} \leq 1, & i \in \{1, 2\}, k \in G, G \in \mathcal{G}, t \in [T].
\end{aligned} \tag{P}$$

2.4. Detalles técnicos de la resolución del problema

2.4.1. Generación de las probabilidades de elección

Para que todo lo anterior tenga sentido es necesario que las probabilidades de elección sean conocidas. Para efectos de la resolución se supone que existe un vector $\hat{\mathbf{u}}$ que genera estas probabilidades dependiendo del cuestionario $(\mathbf{x}_1, \mathbf{x}_2)$, como un oráculo de consulta cada vez que estos números se necesiten. Sin perder generalidad, y así como se hizo en el programa anterior, se impone que $\hat{u}_{1H} = 0$.

2.4.2. Linealizaciones

Para poder trabajar con el problema $(P_{x,y})$ de forma cómoda es necesario realizar dos pasos previos de linealización, pues éste no es lineal.

El primer paso consiste en linealizar la función objetivo, introduciendo una variable w_{kG} que tome el rol del valor absoluto $|z_{kG} - z'_{kG}|$, una variable binaria d_{kG} que vale 0 si y sólo si $z_{kG} - z'_{kG} = w_{kG}$ y una constante K_{kG} lo suficientemente grande, dado que se busca maximizar dicho valor absoluto. Las restricciones que se agregan son

$$\begin{aligned} z_{kG} - z'_{kG} &\leq w_{kG}, & z'_{kG} - z_{kG} &\leq w_{kG}, \\ z_{kG} - z'_{kG} + K_{kG}d_{kG} &\geq w_{kG}, & z'_{kG} - z_{kG} + K_{kG}(1 - d_{kG}) &\geq w_{kG}. \end{aligned}$$

El valor de K_{kG} puede ser elegido notando que $2w_{kG} \leq 2M \leq K_{kG}$, pues $0 \leq z_{kG}, z'_{kG} \leq M$.

El segundo paso consiste en linealizar los productos $z_{kG}y_{itkG}$ y $z'_{kG}y_{itkG}$. Éstos se linealizan introduciendo variables, denotadas s_{itkG} y s'_{itkG} , que representan al producto, suponiendo conocidas cotas inferiores y superiores de z_{kG} . Dichas cotas están fijas de antemano, y valen 0 y M . Por ende, las restricciones que se agregan son

$$\begin{aligned} 0 &\leq s_{itkG} \leq My_{itkG}, & 0 &\leq z_{kG} - s_{itkG} \leq M(1 - y_{itkG}), \\ 0 &\leq s'_{itkG} \leq My_{itkG}, & 0 &\leq z'_{kG} - s'_{itkG} \leq M(1 - y_{itkG}). \end{aligned}$$

Omitiendo los dominios $i \in \{1, 2\}$, $k \in G$, $G \in \mathcal{G}$ y $t \in [T]$, y entendiendo cuál corresponde a qué restricción por contexto, el problema $(P_{x,y})$ linealizado queda como sigue:

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) &= \text{máx} \sum_{G \in \mathcal{G}} \sum_{k \in G} w_{kG} \\ \text{sujeto a} \quad &\sum_{G \in \mathcal{G}} \sum_{k \in G} [p(\mathbf{x}_{2t})s_{1tkG} - p(\mathbf{x}_{1t})s_{2tkG}] = 0, \\ &\sum_{G \in \mathcal{G}} \sum_{k \in G} [p(\mathbf{x}_{2t})(s_{1tkG} - s'_{1tkG}) - p(\mathbf{x}_{1t})(s_{2tkG} - s'_{2tkG})] = 0, \\ &z_{kG} - z'_{kG} \leq w_{kG}, & s_{itkG} &\leq z_{kG}, \\ &z'_{kG} - z_{kG} \leq w_{kG}, & s'_{itkG} &\leq z'_{kG}, \\ &z_{kG} - z'_{kG} + K_{kG}d_{kG} \geq w_{kG}, & z_{1H} &= 1, \\ &z'_{kG} - z_{kG} + K_{kG}(1 - d_{kG}) \geq w_{kG}, & z'_{1H} &= 1, \\ &s_{itkG} \leq My_{itkG}, & \mathbf{z}, \mathbf{z}' &\leq M\mathbf{1}, \\ &s'_{itkG} \leq My_{itkG}, & \mathbf{z}, \mathbf{z}', \mathbf{w}, \mathbf{s}, \mathbf{t} &\geq 0, \\ &z_{kG} - s_{itkG} \leq M(1 - y_{itkG}), & d_{kG} &\in \{0, 1\}. \\ &z'_{kG} - s'_{itkG} \leq M(1 - y_{itkG}), \end{aligned}$$

2.4.3. Número de preguntas

Antes de resolver el problema, se debe considerar el número de preguntas T . La siguiente proposición muestra una cota superior para T , suponiendo una hipótesis en $\hat{\mathbf{u}}$.

Proposición 2.1

Para un conjunto de atributos y niveles \mathcal{G} , sean T_1 y T_2 como sigue:

$$T_1 = \sum_{G \in \mathcal{G}} (n_G - 1) \quad \text{y} \quad T_2 = |\mathcal{G}| - 1.$$

Si el vector $\hat{\mathbf{u}}$ cumple que $\hat{u}_{2G} - \hat{u}_{1G} \neq \hat{u}_{2H}$ para todo G distinto del primer atributo H , entonces existe un diseño con $T^* = T_1 + T_2$ preguntas que, dadas las probabilidades de elección, permite recuperar un sólo vector de valoraciones \mathbf{z} .

Demostración. Se construye un diseño con $T_1 + T_2$ preguntas como se detalla a continuación. Las primeras T_1 preguntas se construyen como sigue: para G un atributo con n_G niveles, una de las alternativas (s.p.g. la primera) queda descrita por el nivel 1 de G , la otra por el nivel $k \in \{2, \dots, n_G\}$, y entre ambas el resto de los atributos $\tilde{G} \neq G$ quedan en el mismo nivel (cualquiera). Haciendo esto para todo $k \in \{2, \dots, n_G\}$ se generan $n_G - 1$ preguntas. Haciendo el mismo proceso para todos los atributos $G \in \mathcal{G}$, se construyen T_1 preguntas. Estas pregunta dependen del nivel k del atributo G , por lo que se denotan t_{kG} .

Las siguientes T_2 preguntas se construyen como sigue: para el par de atributos H y $G \neq H$, la primera alternativa queda descrita por el nivel 1 de H y G , la segunda alternativa por el nivel 2 de H y G , y entre ambas el resto de los atributos $\tilde{G} \neq H, \tilde{G} \neq G$ quedan en el mismo nivel (cualquiera). Haciendo lo mismo para todos los atributos $G \neq H$, se construyen T_2 preguntas. Estas preguntas dependen sólo del atributo G , por lo que se denotan t_G .

Se demuestra ahora que estas preguntas permiten recuperar un único vector \mathbf{z} . En efecto, tomando una pregunta t_{kG} de las primeras T_1 preguntas, la restricción (2.6) correspondiente queda de la siguiente forma:

$$p(\mathbf{x}_{2t_{kG}})z_{1G} - p(\mathbf{x}_{1t_{kG}})z_{kG} = 0 \implies z_{kG} = \frac{p(\mathbf{x}_{2t_{kG}})}{p(\mathbf{x}_{1t_{kG}})}z_{1G}.$$

Nótese que, como las probabilidades están en el intervalo $(0, 1)$, se puede despejar sin problemas. Así, todo z_{kG} depende linealmente del primer nivel de atributo z_{1G} .

Tomando ahora una pregunta t_G de las siguientes T_2 preguntas se tiene que la restricción (2.6) correspondiente queda como sigue:

$$\begin{aligned} 0 &= p(\mathbf{x}_{2t_G})(z_{1H} + z_{1G}) - p(\mathbf{x}_{1t_G})(z_{2H} + z_{2G}) \\ &= p(\mathbf{x}_{2t_G})(z_{1H} + z_{1G}) - p(\mathbf{x}_{1t_G}) \left[\frac{p(\mathbf{x}_{2t_{2H}})}{p(\mathbf{x}_{1t_{2H}})}z_{1H} + \frac{p(\mathbf{x}_{2t_{2G}})}{p(\mathbf{x}_{1t_{2G}})}z_{1G} \right]. \end{aligned}$$

Despejando z_{1G} e imponiendo que $z_{1H} = 1$, se llega a la siguiente expresión:

$$z_{1G} \left[1 - \frac{p(\mathbf{x}_{1t_G}) p(\mathbf{x}_{2t_{2G}})}{p(\mathbf{x}_{2t_G}) p(\mathbf{x}_{1t_{2G}})} \right] = \frac{p(\mathbf{x}_{1t_G}) p(\mathbf{x}_{2t_{2H}})}{p(\mathbf{x}_{2t_G}) p(\mathbf{x}_{1t_{2H}})} - 1. \quad (2.10)$$

Sean $C_1 = \frac{p(\mathbf{x}_{1t_G}) p(\mathbf{x}_{2t_{2G}})}{p(\mathbf{x}_{2t_G}) p(\mathbf{x}_{1t_{2G}})}$ y $C_2 = \frac{p(\mathbf{x}_{1t_G}) p(\mathbf{x}_{2t_{2H}})}{p(\mathbf{x}_{2t_G}) p(\mathbf{x}_{1t_{2H}})}$. Se demuestra ahora que $C_1 < 1 \iff C_2 > 1$.

$$\begin{aligned} C_1 < 1 &\iff \frac{\exp(\hat{u}_{1H}) + \exp(\hat{u}_{1G}) \exp(\hat{u}_{2G})}{\exp(\hat{u}_{2H}) + \exp(\hat{u}_{2G}) \exp(\hat{u}_{1G})} < 1 \\ &\iff \frac{1 + \exp(\hat{u}_{1G}) \exp(\hat{u}_{2G})}{\exp(\hat{u}_{2H}) + \exp(\hat{u}_{2G}) \exp(\hat{u}_{1G})} < 1 \\ &\iff \frac{\exp(\hat{u}_{2G}) + \exp(\hat{u}_{1G} + \hat{u}_{2G})}{\exp(\hat{u}_{1G} + \hat{u}_{2H}) + \exp(\hat{u}_{1G} + \hat{u}_{2G})} < 1 \end{aligned}$$

$$\begin{aligned}
&\iff \exp(\hat{u}_{2G}) + \exp(\hat{u}_{1G} + \hat{u}_{2G}) < \exp(\hat{u}_{1G} + \hat{u}_{2H}) + \exp(\hat{u}_{1G} + \hat{u}_{2G}) \\
&\iff \exp(\hat{u}_{2G}) < \exp(\hat{u}_{1G} + \hat{u}_{2H}) \\
&\iff \exp(\hat{u}_{2H}) + \exp(\hat{u}_{2G}) < \exp(\hat{u}_{2H}) + \exp(\hat{u}_{1G} + \hat{u}_{2H}) \\
&\iff 1 < \frac{\exp(\hat{u}_{2H}) + \exp(\hat{u}_{1G} + \hat{u}_{2H})}{\exp(\hat{u}_{2H}) + \exp(\hat{u}_{2G})} \\
&\iff 1 < \frac{1 + \exp(\hat{u}_{1G})}{\exp(\hat{u}_{2H}) + \exp(\hat{u}_{2G})} \frac{\exp(\hat{u}_{2H})}{1} \\
&\iff 1 < \frac{\exp(\hat{u}_{1H}) + \exp(\hat{u}_{1G})}{\exp(\hat{u}_{2H}) + \exp(\hat{u}_{2G})} \frac{\exp(\hat{u}_{2H})}{\exp(\hat{u}_{1H})} = C_2.
\end{aligned}$$

El mismo argumento sirve para demostrar que $C_1 > 1 \iff C_2 < 1$. Gracias a esta propiedad se tiene que en estos casos z_{1G} queda únicamente determinado y es no negativo, acorde a las restricciones de (P).

Más aún, con la técnica anterior se puede demostrar que $C_1 = 1 \iff C_2 = 1$, pero este caso es problemático. Si $C_1 = 1$ o $C_2 = 1$ entonces la igualdad (2.10) es una del estilo « $0 = 0$ » y el valor de z_{1G} puede ser cualquiera. Para evitar esto se impone la condición $\hat{u}_{2G} - \hat{u}_{1G} \neq \hat{u}_{2H}$. En efecto,

$$\begin{aligned}
C_2 = 1 &\iff \frac{\exp(\hat{u}_{2H}) + \exp(\hat{u}_{1G} + \hat{u}_{2H})}{\exp(\hat{u}_{2H}) + \exp(\hat{u}_{2G})} = 1 \\
&\iff \exp(\hat{u}_{2H}) + \exp(\hat{u}_{1G} + \hat{u}_{2H}) = \exp(\hat{u}_{2H}) + \exp(\hat{u}_{2G}) \\
&\iff \exp(\hat{u}_{1G} + \hat{u}_{2H}) = \exp(\hat{u}_{2G}) \\
&\iff \hat{u}_{1G} + \hat{u}_{2H} = \hat{u}_{2G} \\
&\iff \hat{u}_{2H} = \hat{u}_{2G} - \hat{u}_{1G}.
\end{aligned}$$

Imponiendo la inequidad se tiene que el vector \mathbf{z} queda únicamente determinado. Para demostrar que $\mathbf{z}' = \mathbf{z}$ basta usar el mismo argumento que demuestra que \mathbf{z} queda determinado, pero considerando las restricciones (2.7). \square

Cabe destacar que, en las T_2 preguntas, fijar los niveles de los atributos H y $G \neq H$ en 1 y 2 no es arbitraria. Se puede pensar *a priori* que en las T_2 preguntas cualquier par de niveles sirve. Sin embargo, puede haber un atributo con sólo dos niveles. De ahí que éstos se fijen en 1 y 2.

La restricción sobre el vector $\hat{\mathbf{u}}$ merece un poco de discusión. Si \hat{u}_{1H} tuviera otro valor distinto de 0, entonces la restricción sería en realidad $\hat{u}_{2G} - \hat{u}_{1G} \neq \hat{u}_{2H} - \hat{u}_{1H}$ para todo atributo $G \neq H$. La restricción acota los posibles valores de $\hat{\mathbf{u}}$ en un comienzo, pero asegura que exista un diseño con un número de preguntas razonable en la práctica. Por eso es que se impone la condición. Nótese que $T^* = \sum_{G \in \mathcal{G}} n_G - 1$, lo que se traduce en que el «peor» cuestionario (con más preguntas) tendrá una pregunta por nivel de atributo, salvo el primero (de ahí el término «-1»).

La proposición anterior permite fijar un valor para T con el cual se asegura la existencia de un diseño que identifica al problema. Para efectos de la resolución de (P), se fija $T = T^*$. La figura 2.1 muestra un gráfico con el valor de T^* en función del número de atributos.

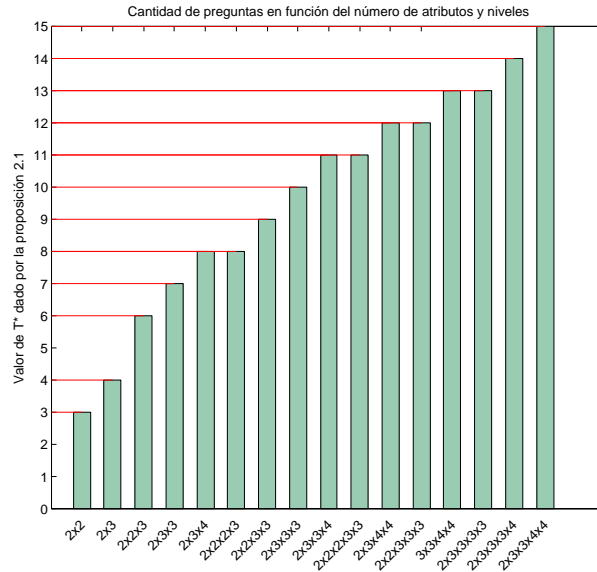


Figura 2.1: Cantidad de preguntas según número de atributos y niveles para el caso de dos alternativas.

2.4.4. Quiebres de simetría

Un tercer detalle a considerar tiene que ver con las simetrías del problema de nivel superior. En un conjunto de elección el orden de las alternativas no es relevante (al menos para este trabajo). Sin embargo, esto no está impuesto como restricción en (P). Esto causa que para todo par de alternativas factibles \mathbf{x}_{1t} , \mathbf{x}_{2t} , las tuplas $(\mathbf{x}_{1t}, \mathbf{x}_{2t})$ y $(\mathbf{x}_{2t}, \mathbf{x}_{1t})$ sean soluciones factibles de (P), lo que hace que la búsqueda del óptimo sea más difícil. En general, si el conjunto de elección tiene N alternativas, entonces para una solución factible $(\mathbf{x}_{1t}, \dots, \mathbf{x}_{Nt})$ y una permutación $\tau \in \Sigma_N$ se tiene que $(\mathbf{x}_{\tau(1)t}, \dots, \mathbf{x}_{\tau(N)t})$ es una solución factible con el mismo valor objetivo. Para solucionar este problema se impone una restricción lexicográfica como sigue:

$$\mathbf{x}_{1t} \leq_L \mathbf{x}_{2t}, \quad t \in [T], \quad (2.11)$$

lo que se traduce en la siguiente desigualdad:

$$\sum_{k=1}^n 2^{n-k} x_{1tk} \leq \sum_{k=1}^n 2^{n-k} x_{2tk}, \quad t \in [T],$$

donde, por simplicidad, se omite la notación « kG » y n es el número de todos los niveles de todos los atributos. Esta restricción es corolario de un teorema en donde se muestra cómo quebrar simetrías de un problema entero usando su grupo de simetría asociado [Mar10].

Análogamente, para efectos de este trabajo el orden de las preguntas tampoco es relevante. Se produce exactamente el mismo problema que antes: dos preguntas son intercambiables y sus codificaciones son factibles para (P). Para solucionar este problema se impone también una restricción lexicográfica como sigue:

$$\begin{pmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \end{pmatrix} \leq_L \begin{pmatrix} \mathbf{x}_{1,t+1} \\ \mathbf{x}_{2,t+1} \end{pmatrix}, \quad t \in \{1, \dots, T-1\}. \quad (2.12)$$

2.4.5. Resolución computacional

Una vez impuestas todas las restricciones anteriores se procede a resolver este problema. Los problemas MIP binivel son difíciles de resolver, por lo que existen algoritmos de aproximación que permiten trabajar con ellos. Para intentar resolver este problema se trabaja con una adaptación de un algoritmo de generación de columnas y restricciones, propuesto por Zeng y An [ZA14].

Como el algoritmo se basa generación de columnas y restricciones, es necesario definir un problema maestro y un subproblema. En este caso, el subproblema es $(P_{x,y})$ y el problema maestro tiene la estructura de (P) . Se resuelve el problema maestro, que entrega un cuestionario. Ese cuestionario se usa como dato para resolver el subproblema. Si el valor óptimo de dicho cuestionario es 0 (o si se alcanza una cierta tolerancia), entonces se rescata ese cuestionario como solución. En caso contrario, se agregan columnas y restricciones al problema maestro para que sea resuelto de nuevo. Más detalles del algoritmo y su adaptación se pueden encontrar en el apéndice B.

Un detalle a destacar tiene que ver con encontrar un cuestionario que no alcance la tolerancia antes mencionada. Si un cuestionario no logra esto, entonces no es necesario considerarlo nuevamente como solución factible. Esto se puede hacer agregando una restricción lineal al problema maestro. En concreto, supóngase que un cuestionario $\tilde{\mathbf{x}}$ (visto como vector) ya fue usado y no se cumplen los criterios de parada del algoritmo. Se puede agregar al problema maestro la restricción

$$\langle \tilde{\mathbf{x}}, \mathbf{x} \rangle = \sum_k \tilde{x}_k x_k \leq 2T|\mathcal{G}| - 1,$$

donde se omiten el subíndice « $itkG$ » por simplicidad. En efecto, esta restricción permite no repetir el cuestionario $\tilde{\mathbf{x}}$, pues si $\mathbf{x} = \tilde{\mathbf{x}}$, entonces $\langle \tilde{\mathbf{x}}, \mathbf{x} \rangle = 2T|\mathcal{G}|$, pues cada vector posee $2T|\mathcal{G}|$ coordenadas iguales a 1.

Se usa esta idea para plantear la verdadera restricción lineal para no repetir cuestionarios, pero en las variables y . Nótese que estos vectores y pueden «ser generados» por varios cuestionarios \mathbf{x} , pues los vectores y sólo reflejan la unicidad de los (niveles de) atributos, ignora aquellos niveles que se repiten, y son éstos los que se usan en $(P_{x,y})$ para identificar. Luego, si se impone en (P) que las variables y no sean iguales a las variables \tilde{y} que ya fueron generadas, se prohíben aún más cuestionarios: todos aquellos que generan a las variables \tilde{y} .

Considerando la discusión anterior, si el problema maestro se ha resuelto ℓ veces e $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(\ell)}$ son los vectores encontrados en cada iteración, entonces se agregan las siguientes restricciones:

$$\langle \mathbf{y}^{(l)}, \mathbf{y} \rangle \leq \sum_k (y_{1k}^l + y_{2k}^l) - 1, \quad l \in \{1, \dots, \ell\}. \quad (2.13)$$

Otro detalle tiene que ver con cómo se inicia la generación de columnas y restricciones. En la primera iteración no hay restricciones más que aquellas que definen al diseño. Además, en la función objetivo estas variables no aparecen. Luego, cualquier diseño es óptimo en la primera iteración. Para salvaguardar esto e inicializar el algoritmo, se escoge un diseño al azar como punto de partida. Las únicas restricciones de este diseño inicial son (1) que tenga T^* conjuntos de elección, con T^* igual al valor dado por la proposición 2.1; y (2) que por cada conjunto de elección las alternativas sean distintas.

El siguiente pseudocódigo resume la resolución de (P) .

Algoritmo 2 Resolución de (P).

Entrada: Contexto abstracto de atributos y niveles, vector $\hat{\mathbf{u}}$ que representa las valoraciones reales de los atributos y que cumpla con la proposición 2.1, punto de partida $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$, tolerancia $\varepsilon > 0$.

Salida: Cuestionario $(\mathbf{x}_1, \mathbf{x}_2)$.

- 1: Definir restricciones del problema maestro (PM) según (P).
 - 2: Definir subproblema (SP) según $(P_{\mathbf{x},\mathbf{y}})$ y la subsección 2.4.2, usando $(\mathbf{x}^{(l)}, \mathbf{y}^{(l)})$ y $\hat{\mathbf{u}}$ para generar los coeficientes.
 - 3: $LB = -\infty, UB = \infty, l = 0$, resolver (SP).
 - 4: **si** $f(\mathbf{x}, \mathbf{y}) = 0$ **entonces devolver** $(\mathbf{x}^{(l)}, \mathbf{y}^{(l)}, \mathbf{z}^*, \mathbf{w}^*)$
 - 5: **si no**
 - 6: $l \leftarrow l + 1, UB \leftarrow \min\{UB, \text{valor de (SP)}\}$
 - 7: Generar columnas y restricciones en (PM) usando coeficientes de (SP). Generar restricciones (2.11), (2.12), (2.13) en (PM). Resolver (PM). $LB = \text{valor de (PM)}$.
 - 8: **si** $UB - LB \leq \varepsilon$ **entonces devolver** UB y solución incumbente. Terminar.
 - 9: **si no**
 - 10: Tomar solución de (PM) $(\mathbf{x}^{(l)}, \mathbf{y}^{(l)})$ y usarla para generar y resolver (SP). Ir a 4.
 - 11: **fin si**
 - 12: **fin si**
-

Dado lo anterior, para una instancia fija de atributos y niveles se escogen diez puntos de partida distintos y se resuelve (P) diez veces. El siguiente pseudocódigo resume el procedimiento a seguir para obtener los cuestionarios.

Algoritmo 3 Procedimiento de resolución.

- 1: Fijar un contexto (abstracto) de atributos y niveles.
 - 2: Generar vector $\hat{\mathbf{u}}$ (valoraciones reales de los atributos y que cumpla con la proposición 2.1).
 - 3: Generar diez cuestionarios al azar, distintos, $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(10)})$.
 - 4: **para** $i = 1$ hasta 10 **hacer**
 - 5: Resolver (P) considerando lo discutido en las secciones 2.4.2 – 2.4.5, con $\mathbf{x}^{(i)}$ como punto de partida. Guardar la solución $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*, \mathbf{w}^*)$ de (P).
 - 6: **fin para**
-

Los contextos abstractos de atributos y niveles que se fijan para estudiar son las instancias 2^2 , 2×3 , $2^2 \times 3$, 2×3^2 , $2 \times 3 \times 4$, $2^3 \times 3$ y $2^2 \times 3^2$, cada una con un vector $\hat{\mathbf{u}}$. Cabe destacar que, una vez obtenido un diseño, se preprocesa para eliminar conjuntos de elección o bien repetidos o bien con alternativas repetidas. Ejemplos de diseños para el modelo EBA y compensatorio se dan en el apéndice D.

2.5. Resultados y análisis

En esta sección se contrastan los diseños arrojados por el algoritmo propuesto *versus* diseños compensatorios clásicos, además de propiedades inherentes del diseño. La estrategia a seguir es la siguiente: para cada cuestionario EBA se simulan respuestas con el vector $\hat{\mathbf{u}}$ dado en la sección 2.4,

y con dichas respuestas se estiman los parámetros con la técnica de máxima verosimilitud. Luego se calculan métricas de optimalidad para comparar diseños EBA contra diseños compensatorios clásicos. Por último, se analizan características descriptivas de los diseños.

El estudio computacional se realiza en un computador portátil (con cuatro procesadores a 4,2 GHz y 8,00 GB de memoria RAM), con IBM ILOG CPLEX (versión 12.6.2) integrado con MATLAB (versión R2013a) como *solver* MIP. Se fija la tolerancia del problema maestro y de todo el algoritmo a 1 %, la de los subproblemas a 0 % y el tiempo límite de computación a 1200 segundos.

2.5.1. Tiempo de ejecución

La tabla 2.1 muestra, para cada instancia, el número de variables enteras, el tiempo promedio de resolución de las diez veces que se resuelve el problema, y el tiempo máximo de esas diez veces. Se puede apreciar que en promedio el algoritmo corre bastante rápido y, como es de esperar, los tiempos de resolución son más altos cuando el número de variables enteras aumenta. Si bien esto hace sugerir que el problema no es escalable, en la práctica estos tiempos de resolución son pequeños. Sin, ocurre en contadas ocasiones que el punto de partida no permite que la resolución sea rápida, llegando al límite de los 1200 segundos.

Tabla 2.1: Número de variables y tiempo de resolución para distintas instancias.

Instancia	Número de variables enteras	Tiempo promedio de resolución [s]	Máximo tiempo de ejecución [s]
2^2	24	0,3929	0,5072
2×3	40	1,1041	1,4995
$2^2 \times 3$	84	18,7657 (127,2689) ¹	1200
2×3^2	112	20,9126	30,0848
$2 \times 3 \times 4$	144	22,0941	30,2486
$2^3 \times 3$	144	33,7224 (369,7120) ¹	1200
$2^2 \times 3^2$	180	30,8191 (134,6419) ¹	1200

1. El primer valor señala el promedio entre las instancias que no llegaron al límite de 1200 segundos; el valor entre paréntesis señala el promedio de las diez instancias.

2.5.2. Tamaño del diseño

Una primera pregunta que surge con respecto a los diseños obtenidos por el programa es ¿qué tan ajustado es el valor de T^* ? La tabla 2.2 muestra, por contexto, los valores de T^* dado por la proposición 2.1, además del número de preguntas del diseño más pequeño, del más grande y el valor promedio entre los diez obtenidos.

Una primera interpretación de estos resultados es que, si bien en todos los contextos se obtuvo un cuestionario con el número de preguntas igual a T^* , éste no es un valor muy restrictivo. Sin embargo, esto no es decidor para concluir cuál es el cuestionario a ocupar para recuperar los parámetros.

Tabla 2.2: Comparación entre el número de preguntas para distintas instancias.

Instancia	T^* (prop. 2.1)	Mínimo T	Máximo T	Promedio de T
2^2	3	2	3	2,1
2×3	4	2	4	3,9
$2^2 \times 3$	6	3	6	4,8
2×3^2	7	4	7	5,6
$2 \times 3 \times 4$	8	4	8	6,2
$2^3 \times 3$	8	5	8	6,3
$2^2 \times 3^2$	9	6	9	7,3

2.5.3. D -eficiencia y características descriptivas

Como se menciona en la sección 1.5, una forma de medir qué tan óptimo es un diseño consiste en analizar la matriz de información de Fisher del modelo. A grandes rasgos, esta matriz es la inversa de la matriz de varianzas-covarianzas asintótica, por lo que «minimizar la varianza» y «maximizar la información» son similares. Como son matrices, se trabaja con funciones a valores reales de las matrices.

De las tantas medidas de optimalidad se escoge la D -eficiencia (donde la función a considerar es el determinante) por ser clásica en el campo. La matriz de información de Fisher de este modelo particular y sus propiedades pueden verse en el apéndice C, sección C.1.

La metodología para calcular la D -eficiencia es la siguiente para cada cuestionario: se considera el cuestionario \mathbf{x} y el vector \mathbf{z} entregados por la resolución del problema (P), se maximiza la log-verosimilitud del modelo (C.1) con \mathbf{z} como punto de partida, obteniendo un estimador de los pesos del modelo $\hat{\alpha}_{\mathbf{x}}$, y se calcula la matriz de información de Fisher del modelo y su determinante normalizado:

$$D^{\text{EBA-error}} = \det[\mathcal{I}(\hat{\alpha}_{\mathbf{x}})]^{-1/n},$$

donde n es el número de parámetros a estimar. Cabe destacar que la matriz de información difiere sustancialmente a la matriz del modelo compensatorio. Esta diferencia lleva a diseños muy distintos. A modo de ejemplo, la tabla 2.3 muestra, para los diez cuestionarios obtenidos con el algoritmo propuesto en un contexto fijo, su número de preguntas y su $D^{\text{EBA-error}}$.

Se puede apreciar que no existe una relación directa entre el número de preguntas del cuestionario y su $D^{\text{EBA-error}}$ calculado. Si bien tener más preguntas permite comparar y diferenciar distintas combinaciones de pesos, esto no se ve reflejado en la eficiencia del cuestionario (en este ejemplo, el cuestionario óptimo es el número 9, con seis preguntas). La misma tabla 2.3 muestra que cinco diseños, todos con seis preguntas cada uno, tiene asociados distintos valores de $D^{\text{EBA-error}}$.

Tabla 2.3: D^{EBA} -error para el caso $2 \times 3 \times 3$ versus número de preguntas.

Número de cuestionario	Número de preguntas	D^{EBA} -error
1	5	$2,4904 \times 10^{-23}$
2	7	$1,1063 \times 10^{-23}$
3	5	$1,6408 \times 10^{-23}$
4	5	34,1755
5	6	$7,6333 \times 10^{-24}$
6	4	$3,7177 \times 10^{-21}$
7	6	12,4985
8	6	13,3551
9	6	$1,6407 \times 10^{-26}$
10	6	$2,9537 \times 10^{-23}$

La tabla 2.4 resume, para cada contexto de niveles y atributos, el D^{EBA} -error más pequeño encontrado entre los diez cuestionarios, su número de preguntas y el D^{EBA} -error asociado al cuestionario dado por la proposición 2.1. Comparando los dos valores mostrados, no se puede obtener una conclusión decidora con respecto a qué tanto mejor es el cuestionario dado en la proposición. En algunos contextos de atributos este cuestionario resultó ser mejor (en el sentido de la D -eficiencia) que el mejor cuestionario arrojado por la resolución de (P), pero en otros casos sucede lo contrario.

Tabla 2.4: Datos de los mejores cuestionarios EBA por contexto.

Contexto	Número de preguntas	Mejor D^{EBA} -error	Número de preguntas (prop 2.1)	D^{EBA} -error (prop 2.1)
2^2	2	$6,6135 \times 10^{-16}$	3	$3,4908 \times 10^{-16}$
2×3	4	$3,2125 \times 10^{-19}$	4	$2,1670 \times 10^{-19}$
$2^2 \times 3$	5	$1,4096 \times 10^{-22}$	6	$7,9270 \times 10^{-23}$
2×3^2	6	$1,6407 \times 10^{-26}$	7	$5,2192 \times 10^{-24}$
$2 \times 3 \times 4$	7	$3,9085 \times 10^{-25}$	8	$1,4411 \times 10^{-24}$
$2^3 \times 3$	6	$1,8562 \times 10^{-24}$	8	$1,1393 \times 10^{-24}$
$2^2 \times 3^2$	9	$1,9008 \times 10^{-25}$	9	$2,3287 \times 10^{-25}$

Pérdidas de eficiencia dado un modelo subyacente

En la ausencia de direcciones para diseñar un experimento para el modelo EBA, los diseños compensatorios son la alternativa práctica, pues ya hay toda una teoría desarrollada al respecto. Se estudia a continuación el desempeño de los diseños compensatorios cuando el modelo subyacente es EBA. Para el mejor diseño compensatorio \mathbf{x} se mide su pérdida de eficiencia como

$$\text{Pérdida de eficiencia basada en el } D^{\text{EBA}}\text{-error}(\mathbf{x}) = 1 - \frac{D^{\text{EBA}}\text{-error}(\mathbf{x}^*)}{D^{\text{EBA}}\text{-error}(\mathbf{x})}, \quad (2.14)$$

donde \mathbf{x}^* es el diseño D^{EBA} -óptimo de la búsqueda computacional.

Asimismo se puede plantear la medida conversas: qué tan eficiente es un diseño EBA cuando el modelo subyacente es compensatorio. Se utiliza la misma forma anterior, con la excepción de que ahora la pérdida se mide con respecto al $D^{\text{Comp-error}}$, obtenido en base al determinante de otra matriz de información, calculada para un modelo logit clásico [HZ96]. Para el mejor diseño EBA \mathbf{x} se mide su pérdida de eficiencia como

$$\text{Pérdida de eficiencia basada en el } D^{\text{Comp-error}}(\mathbf{x}) = 1 - \frac{D^{\text{Comp-error}}(\mathbf{x}^*)}{D^{\text{Comp-error}}(\mathbf{x})},$$

donde \mathbf{x}^* es el diseño D^{comp} -óptimo de la búsqueda computacional.

Una tercera medida de pérdida de eficiencia se basa en el error cuadrático medio (RMSE por sus siglas en inglés) del estimador $\hat{\alpha}$. En particular, si se denota $\hat{\alpha}_{\mathbf{x}}$ al estimador obtenido usando el diseño \mathbf{x} , entonces se define

$$\text{RMSE}(\hat{\alpha}_{\mathbf{x}}) = \sqrt{\frac{1}{n} \sum_{k=1}^n (\alpha_k - \hat{\alpha}_k)^2},$$

donde α es el verdadero vector de pesos. La pérdida de eficiencia basada en la recuperación de los parámetros se define similar a como se hizo en (2.14) reemplazando el $D^{\text{EBA-error}}$ por el RMSE correspondiente. Si bien no es explícito, la estimación depende del número de datos (ver apéndice C para más detalles), por lo que para un diseño dado se simulan mil respuestas basadas en el modelo subyacente y se obtienen el RMSE correspondiente.

Tabla 2.5: Pérdida de eficiencia según distintas métricas.

(a) El modelo subyacente es EBA			(b) El modelo subyacente es compensatorio		
Contexto	Porcentaje de pérdida de eficiencia basada en		Contexto	Porcentaje de pérdida de eficiencia basada en	
	$D^{\text{EBA-error}}$	RMSE		$D^{\text{Comp-error}}$	RMSE
2^2	100 ¹	0,5336	2^2	40,7359	53,7044
2×3	100	0,2109	2×3	-	-
$2^2 \times 3$	100 ¹	0,3897	$2^2 \times 3$	46,6164	16,2185
2×3^2	100 ¹	0 (-0,1376) ²	2×3^2	-	-
$2 \times 3 \times 4$	100 ¹	0 (-0,7640) ²	$2 \times 3 \times 4$	28,4038	70,5398
$2^3 \times 3$	100	0 (-0,0792) ²	$2^3 \times 3$	57,4703	59,2224
$2^2 \times 3^2$	100 ¹	0 (-11,4119) ²	$2^2 \times 3^2$	-	-

1. El valor es cercano a 100, pero por errores de máquina se imprime 100. 2. Valor entre paréntesis: pérdida de eficiencia efectiva según la fórmula. Como el número es negativo, la pérdida de eficiencia es 0.

Para cada contexto se reportan las medidas de pérdida de eficiencia en la tabla 2.5. Cuando el modelo subyacente es EBA (tabla 2.5a), se concluye que en todos los casos los diseños compensatorios clásicos presentan una alta pérdida de eficiencia cuando la métrica base es el $D^{\text{EBA-error}}$.

Por otro lado, las pérdidas de eficiencia basadas en el RMSE son variadas, dependiendo del contexto: en los contextos con menos niveles la pérdida de eficiencia es muy baja, y en los contextos más grandes prácticamente no hay pérdida de eficiencia.

Por otro lado, si el modelo subyacente es compensatorio (tabla 2.5b), se concluye que los diseños EBA que logran identificar el modelo tienen una menor pérdida de eficiencia si la métrica base es el D^{Comp} -error, pero una mayor pérdida de eficiencia si la métrica usada es RMSE. En algunos casos ocurre que el mejor cuestionario EBA no es capaz de identificar el modelo compensatorio subyacente, pues éstos no realizan las suficientes comparaciones entre atributos.

Los resultados finales de esta sección indican datos algo contradictorios. Dependiendo de la métrica usada, los diseños EBA resultan ser mejores o peores que los compensatorios a la hora de estimar los parámetros del modelo. Si se usa la métrica usual (D -eficiencia) como medida de optimalidad, entonces en este caso es más seguro usar diseños EBA para estimar los modelos. Si, por el contrario, se piensa en una métrica de distancia entre los parámetros reales y estimados, entonces los diseños compensatorios parecen ser óptimos a la hora de estimar.

2.5.4. Características descriptivas

Investigaciones existentes en diseños compensatorios enfatizan los siguientes aspectos: balance de niveles, ortogonalidad y traslape de niveles. Se estudia a continuación cómo varían los diseños del modelo EBA de los diseños compensatorios en estos aspectos [Kuh10, LA11].

El balance de niveles mide qué tan a menudo ocurren los niveles de un atributo. Un diseño se dice balanceado si cada nivel ocurre con igual frecuencia. Por ejemplo, cada nivel de un atributo de tres niveles debe aparecer en un tercio de los casos. Si bien esta propiedad es binaria, se puede plantear una métrica de «qué tan balanceado» es un diseño, midiendo la desviación absoluta promedio (DAP) entre la frecuencia esperada de cada nivel y la real, es decir,

$$\text{DAP}(\mathbf{x}) = \frac{1}{n} \sum_{\text{niveles } k} \left(m(k) - \frac{T(\mathbf{x})}{n_{G(k)}} \right),$$

donde $m(k)$ es la cantidad de veces que aparece el nivel k en todas las alternativas y preguntas, $T(\mathbf{x})$ es el número de preguntas de \mathbf{x} y $G(k)$ es el atributo correspondiente al nivel k . Un diseño balanceado tiene una DAP igual a cero, y mientras más grande la DAP, menos balanceado es el diseño. Esta propiedad es deseable en los modelos lineales pues asegura que el intercepto del estimador sea ortogonal a cada efecto.

La ortogonalidad se refiere a correlaciones entre estimadores. Un diseño se dice ortogonal si cada par de niveles aparece con la misma frecuencia entre todos los pares. Así como el balance, esta propiedad es binaria y se puede plantear una métrica de «qué tan ortogonal» es un diseño, midiendo la correlación promedio de a pares de las columnas (ajustadas a la media) de la matriz del diseño, es decir,

$$\text{ORT}(\mathbf{x}) = \frac{1}{\frac{n(n-1)}{2}} \sum_{k=1}^n \sum_{l=k+1}^n \text{corr}(\mathbf{X}_k, \mathbf{X}_l),$$

donde \mathbf{X} es una matriz que codifica al diseño \mathbf{x} [Kuh10, capítulo 3], \mathbf{X}_k es la fila k -ésima de \mathbf{X} y $\text{corr}(\mathbf{X}, \mathbf{Y})$ es la correlación entre \mathbf{X} e \mathbf{Y} . Un diseño ortogonal tiene correlación igual a cero, y mientras más grande el número, menos ortogonal es el diseño. Esta propiedad es deseable

en los diseños de modelos compensatorios porque asegura que los efectos estimables no estén correlacionados.

El traslape de niveles concierne a los conjuntos de elección del diseño y a cuántas veces aparece un nivel de atributo en un conjunto, es decir,

$$TN(\mathbf{x}) = 1 - \frac{1}{T(\mathbf{x})|\mathcal{G}|} \sum_{t=1}^{T(\mathbf{x})} \sum_{G \in \mathcal{G}} c(G, t),$$

donde $c(G, t)$ vale 1 si en la pregunta t los niveles de G que aparecen en las alternativas no se repiten y 0 en otro caso². En los modelos lineales se busca que el traslape de niveles sea mínimo, pues de lo contrario se provee menos información a la hora de estimar los parámetros.

Las figuras 2.2, 2.3 y 2.4 resumen los distintos cálculos de las métricas antes mencionadas. Para cada contexto de atributo se calculan las métricas para los diez diseños arrojados tanto por el algoritmo propuesto como para diseños compensatorios obtenidos vía SPSS (versión 21.0.0.0) [IBM12], para luego promediar dichos valores. Es este promedio el que se muestra en las figuras.

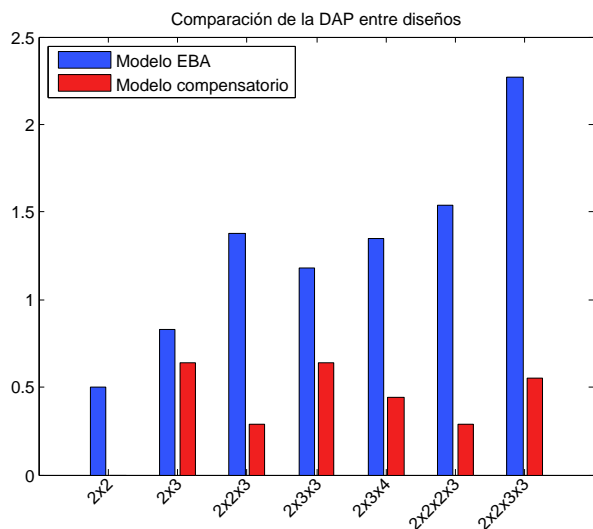


Figura 2.2: Comparación de la DAP.

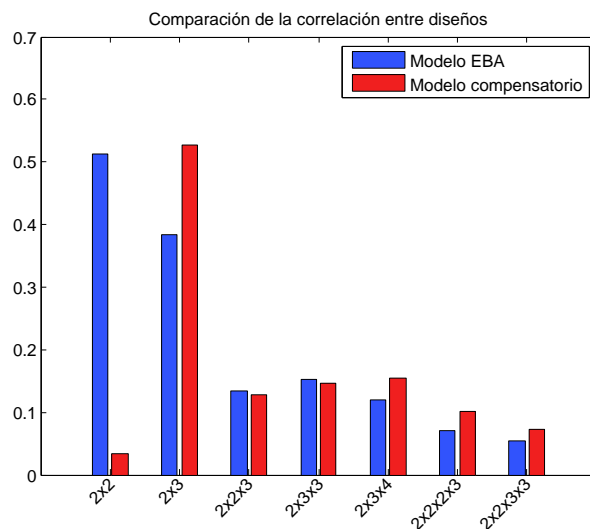


Figura 2.3: Comparación de la correlación.

Una primera conclusión a obtener es que los diseños EBA en promedio presentan menor balance de niveles y mayor traslape de niveles que sus contrapartes compensatorias. Los niveles de ortogonalidad se comportan de forma similar, salvo en los contextos 2×2 y 2×3 , en donde se observan comportamientos extremos.

Existe una muy buena razón para que los diseños EBA posean un nivel de traslape más alto que los diseños compensatorios correspondientes. Si existe un nivel que se traslapa entre dos alternativas, entonces éste no participa de las probabilidades de elección, por lo que son menos los parámetros que actúan en esa pregunta. El nivel de traslape permite así una mejor estimación de los parámetros en el caso EBA.

²En la situación en que el número de niveles sea menor al de alternativas, $c(G, t)$ vale 1 si aparecen todos los niveles, y 0 en otro caso.

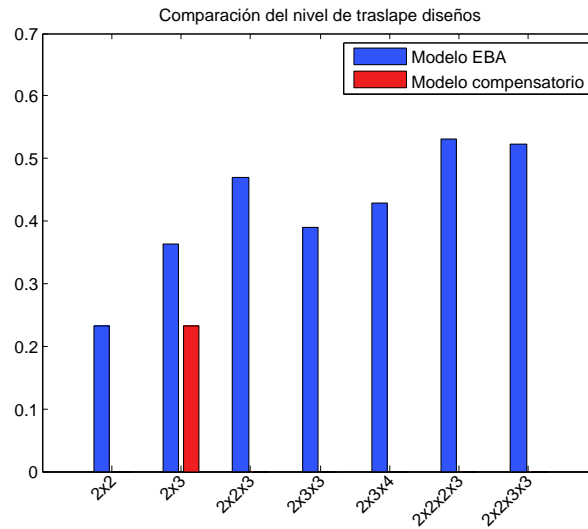


Figura 2.4: Comparación del nivel de traslape.

El nivel de ortogonalidad es muy importante en los modelos compensatorios por lo que se menciona antes (a mayor ortogonalidad, menor la correlación entre los estimadores). Como el modelo EBA no es compensatorio, entonces *a priori* los diseños no tienen por qué presentar esta propiedad, y en efecto ocurre que los diseños están lejos de tenerla.

El balance de niveles tampoco parece ser importantes en los diseños EBA. Esto puede explicarse por lo siguiente: en (P) no se impone cuántas veces han de aparecer los niveles de atributos. Así, puede ocurrir que un nivel k aparezca muchas veces, y todos los otros niveles queden en función de k , así como ocurre en el diseño construido en la proposición 2.1.

Capítulo 3

Un algoritmo de optimización para el desarrollo de cuestionarios con tres alternativas

Se estudia ahora el caso en donde los conjuntos de elección poseen tres alternativas. La estrategia de estudio es análoga a la anterior: se modela el formato de las preguntas como vectores binarios, se explicitan las probabilidades de elección y se plantea y resuelve un problema MIP binivel. Las codificaciones de los atributos y sus niveles no cambia.

3.1. Formato de las preguntas

Siguiendo con la notación anterior, sean \mathbf{x}_{1t} , \mathbf{x}_{2t} y \mathbf{x}_{3t} los perfiles de las tres alternativas del conjunto de elección mostrado en la pregunta $t \in [T]$. Cada componente binaria de \mathbf{x}_{it} se indexa como sigue:

$$x_{itkG} = \begin{cases} 1 & \text{si el aspecto } k \in G \ (G \in \mathcal{G}) \text{ está en la alternativa } i \text{ en la pregunta } t, \\ 0 & \text{en otro caso.} \end{cases} \quad (3.1)$$

Análogo al caso anterior, sea $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})$ para $i \in \{1, 2, 3\}$. Se denota por $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ al diseño que consta de T preguntas y de las tres alternativas mostradas en cada pregunta.

3.2. Las probabilidades

Se analizan ahora las probabilidades de elección. La gran diferencia con el caso de dos alternativas es que en esta ocasión los aspectos que dos alternativas cualquiera comparten *si son relevantes* a la hora de escoger, pero no así los aspectos que están presentes en las tres alternativas a la vez. Sea como antes $A_{it} \subseteq A$ el conjunto de aspectos que aparecen únicamente en la alternativa i en la pregunta t , y sea $A_{ijt} \subseteq A$ el conjunto de aspectos comunes sólo entre las alternativas $i \neq j$ en la pregunta t , con $i, j \in \{1, 2, 3\}$ y $t \in [T]$. Denotando como antes $z_{kG} = \exp(u_{kG})$, y definiendo

$$Z_{3t} = \sum_{k \in A_{1t}} z_{kG} + \sum_{k \in A_{2t}} z_{kG} + \sum_{k \in A_{3t}} z_{kG} + \sum_{k \in A_{12t}} z_{kG} + \sum_{k \in A_{13t}} z_{kG} + \sum_{k \in A_{23t}} z_{kG},$$

entonces las probabilidades de elección son

$$p(\mathbf{x}_{it}) = \frac{1}{Z_{3t}} \left(\sum_{k \in A_{it}} z_{kG} + \sum_{k \in A_{ijt}} z_{kG} p_{ij}(\mathbf{x}_{it}) + \sum_{k \in A_{ilt}} z_{kG} p_{il}(\mathbf{x}_{it}) \right), \quad (3.2)$$

donde $i, j, l \in \{1, 2, 3\}$ son todos distintos entre sí y $p_{ij}(\mathbf{x}_{it})$ es la probabilidad de escoger la alternativa i en la pregunta t dado el conjunto de elección $\{i, j\}$, $i \neq j$:

$$p_{ij}(\mathbf{x}_{it}) = \begin{cases} \frac{\sum_{k \in A_{it} \cup A_{ilt}} z_{kG}}{Z_{3t} - \sum_{k \in A_{ijt} \cup A_{lt}} z_{kG}}, & \text{si } \mathbf{x}_{it} \neq \mathbf{x}_{jt}, \\ \frac{1}{2}, & \text{si } \mathbf{x}_{it} = \mathbf{x}_{jt}. \end{cases} \quad (3.3)$$

Se hace la siguiente distinción: dos alternativas i, j pueden ser distintas, pero sus descripciones $\mathbf{x}_{it}, \mathbf{x}_{jt}$ pueden ser iguales, dependiendo de qué modelen los atributos y sus niveles.

De (3.2), el primer término corresponde a haber elegido un aspecto único de la alternativa i , mientras que los otros dos corresponden a haber elegido primero un aspecto común sólo entre i y j (resp. l) y luego escoger o bien un aspecto único de i o uno común sólo entre i y l (resp. j) de entre los disponibles.

Nótese que este sistema es lineal en z_{kG} sólo si se consideran conocidas las probabilidades de elección $p(\mathbf{x}_{it})$ y $p_{ij}(\mathbf{x}_{it})$. De (3.2) se desprenden un sistema de tres igualdades que son linealmente dependientes. Se consideran sólo las primeras dos, para todo $t \in [T]$:

$$0 = (1 - p(\mathbf{x}_{1t})) \sum_{k \in A_{1t}} z_{kG} - p(\mathbf{x}_{1t}) \sum_{k \in A_{2t} \cup A_{3t}} z_{kG} + (p_{12}(\mathbf{x}_{1t}) - p(\mathbf{x}_{1t})) \sum_{k \in A_{12t}} z_{kG} + \dots \\ \dots + (p_{13}(\mathbf{x}_{1t}) - p(\mathbf{x}_{1t})) \sum_{k \in A_{13t}} z_{kG} - p(\mathbf{x}_{1t}) \sum_{k \in A_{23t}} z_{kG}, \quad (3.4)$$

$$0 = (1 - p(\mathbf{x}_{2t})) \sum_{k \in A_{2t}} z_{kG} - p(\mathbf{x}_{2t}) \sum_{k \in A_{1t} \cup A_{3t}} z_{kG} + (p_{12}(\mathbf{x}_{2t}) - p(\mathbf{x}_{2t})) \sum_{k \in A_{12t}} z_{kG} + \dots \\ \dots + (p_{23}(\mathbf{x}_{2t}) - p(\mathbf{x}_{2t})) \sum_{k \in A_{23t}} z_{kG} - p(\mathbf{x}_{2t}) \sum_{k \in A_{13t}} z_{kG}. \quad (3.5)$$

Se tienen así $2T$ ecuaciones lineales para el vector \mathbf{z} .

Dado que las probabilidades de elección dependen de los aspectos únicos de cada alternativa y de aspectos compartidos entre dos alternativas, es necesario construir a partir del diseño $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ ciertos vectores binarios $(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_{12}, \mathbf{y}_{13}, \mathbf{y}_{23})$ que reflejen estas características, es decir

$$y_{itkG} = \begin{cases} 1 & \text{si el aspecto } k \in G (G \in \mathcal{G}) \text{ está únicamente en la alternativa } i \text{ en la pregunta } t, \\ 0 & \text{en otro caso,} \end{cases}$$

$$y_{ijtkG} = \begin{cases} 1 & \text{si el aspecto } k \in G (G \in \mathcal{G}) \text{ está sólo en las alternativas } i, j \text{ en la pregunta } t, \\ 0 & \text{en otro caso.} \end{cases}$$

La construcción de estos vectores es similar al caso anterior. Sólo se muestran las relaciones lógicas, y su equivalencia en restricciones lineales, de los vectores $\mathbf{y}_1, \mathbf{y}_{12}$. El resto de las relaciones

es análogo, cambiando los roles de 1, 2 y 3.

$$\begin{aligned}
(x_{1tkG} = 0 \implies y_{1tkG} = 0) &\iff y_{1tkG} \leq x_{1tkG}, \\
(x_{1tkG} = 1 \wedge [x_{2tkG} = 0 \wedge x_{3tkG} = 0] \implies y_{1tkG} = 1) &\iff y_{1tkG} \geq x_{1tkG} - x_{2tkG} - x_{3tkG}, \\
(x_{1tkG} = 1 \wedge [x_{2tkG} = 1 \vee x_{3tkG} = 1] \implies y_{1tkG} = 0) &\iff y_{1tkG} \leq 2 - x_{1tkG} - \frac{x_{2tkG} + x_{3tkG}}{2}, \\
(x_{1tkG} = 0 \implies y_{12tkG} = 0) &\iff y_{12tkG} \leq x_{1tkG}, \\
(x_{2tkG} = 0 \implies y_{12tkG} = 0) &\iff y_{12tkG} \leq x_{2tkG}, \\
(x_{1tkG} = 1 = x_{2tkG} \wedge x_{3tkG} = 0 \implies y_{12tkG} = 1) &\iff y_{12tkG} \geq x_{1tkG} + x_{2tkG} - x_{3tkG} - 1, \\
(x_{1tkG} = 1 = x_{2tkG} \wedge x_{3tkG} = 1 \implies y_{12tkG} = 0) &\iff y_{12tkG} \leq 3 - x_{1tkG} - x_{2tkG} - x_{3tkG}.
\end{aligned}$$

Como antes, las variables y_{itkG}, y_{ijtkG} no necesitan ser binarias en la programación, pues las variables x_{itkG} fuerzan a que éstas tomen valores en $\{0, 1\}$. Por ende, se declaran como continuas en $[0, 1]$.

3.3. El problema MIP binivel

El problema a resolver es similar al problema (P) planteado en la sección 2.3, con la siguiente salvedad: en lugar de usar (2.4) como restricciones se usan (3.4) y (3.5). La técnica usada para generar (2.6) y (2.7) también se implementa en este caso: un vector, sin perder generalidad \mathbf{z} , ha de cumplir (3.4) y (3.5) y ambos vectores \mathbf{z}, \mathbf{z}' en conjunto han de cumplir también las restricciones.

Si bien no se muestra acá, y abusando un poco de notación, en el resto de este capítulo se denota por (P) al problema de optimización para encontrar cuestionarios con tres alternativas por pregunta.

3.4. Detalles técnicos de la resolución del problema

La mayoría de los detalles técnicos que se discuten en la sección 2.4 se implementan exactamente igual en este caso. Específicamente, las ideas sobre la generación de las probabilidades de elección, las linealizaciones, los quiebres de simetría, la repetición de cuestionarios y la resolución computacional se aplican de forma análoga a este caso.

3.4.1. Número de preguntas

Así como en el caso de dos alternativas, es necesario imponer un valor a T para efectos de la programación. El siguiente resultado es similar a la proposición 2.1, pero para el caso de tres alternativas. Sin embargo, se imponen más restricciones sobre el vector $\hat{\mathbf{u}}$.

Proposición 3.1

Para un conjunto de atributos y niveles \mathcal{G} , sean T_1 y T_2 como sigue:

$$T_1 = |\{G : n_G = 2\}| + \sum_{\substack{G \in \mathcal{G} \\ n_G > 2}} (n_G - 2) \quad \text{y} \quad T_2 = |\mathcal{G}| - 1.$$

Si el vector $\hat{\mathbf{u}}$ cumple que $\hat{u}_{2G} - \hat{u}_{1G} \neq \hat{u}_{2H}$ (y, en particular, $\hat{u}_{2G} - \hat{u}_{1G} < \hat{u}_{2H}$ si $n_H = 2$) para todo G distinto del primer atributo H , y $\hat{u}_{3G} - \hat{u}_{1G} \neq \hat{u}_{3H}$ cuando $n_H \geq 3$, entonces existe un diseño con $T^* = T_1 + T_2$ preguntas que, dadas las probabilidades de elección, permite recuperar un sólo vector de valoraciones \mathbf{z} .

Demostración. Se construye un diseño con $T_1 + T_2$ preguntas como se detalla a continuación. Las primeras T_1 preguntas se construyen como sigue: para un atributo G , la primera alternativa tiene dicho atributo en el nivel 1, la segunda en el nivel 2 y la tercera en el nivel $k \in \{3, \dots, n_G\}$, pero en el caso en que G posea sólo dos niveles, se impone que la tercera alternativa sea igual a la primera. El resto de los atributos quedan en el mismo nivel (cualquiera) para ambas alternativas. Se generan así $n_G - 2$ cuando G tiene tres niveles y una pregunta cuando G tiene dos niveles. Haciendo lo mismo para todos los atributos $G \in \mathcal{G}$, se construyen T_1 preguntas. Estas pregunta dependen del nivel k del atributo G , por lo que se denotan t_{kG} .

Las siguientes T_2 preguntas se construyen para todo par de atributos H y $G \neq H$, considerando los siguientes casos:

1. $n_H = 2 = n_G$: la primera alternativa tiene los atributos en nivel 1, la segunda en nivel 2 y la tercera en nivel 1;
2. $n_H = 2, n_G \geq 3$: la primera alternativa tiene los atributos en nivel 1, la segunda en nivel 2 y la tercera tiene a H en nivel 1 y a G en nivel 3;
3. $n_H, n_G \geq 3$: la primera alternativa tiene los atributos en nivel 1, la segunda en nivel 2 y la tercera en nivel 3; y
4. $n_H \geq 3, n_G = 2$: se pueden, sin perder generalidad, reordenar los atributos para que H sea aquel con menos niveles, y se reduce a un caso anterior.

Esto genera una pregunta por caso. Haciendo lo mismo para todos los atributos $G \neq H$, se construyen T_2 preguntas que dependen sólo del atributo G , por lo que se denotan t_G .

Se demuestra ahora que estas preguntas permiten recuperar un único vector \mathbf{z} . En efecto, tomando una pregunta t_{kG} de las primeras T_1 preguntas para un atributo G con $n_G \geq 3$, la restricción (3.4) queda de la siguiente forma:

$$(1 - p(\mathbf{x}_{1t_{kG}}))z_{1G} - p(\mathbf{x}_{1t_{kG}})(z_{2G} + z_{kG}) = 0 \implies z_{kG} = \frac{1 - p(\mathbf{x}_{1t_{kG}})}{p(\mathbf{x}_{1t_{kG}})}z_{1G} - z_{2G},$$

mientras que la restricción (3.5) queda

$$(1 - p(\mathbf{x}_{2t_{kG}}))z_{2G} - p(\mathbf{x}_{2t_{kG}})(z_{1G} + z_{kG}) = 0 \implies z_{2G} = \frac{p(\mathbf{x}_{2t_{kG}})}{p(\mathbf{x}_{1t_{kG}})}z_{1G}, \quad (3.6)$$

en donde la implicación se tiene al reemplazar el valor de z_{3G} y despejar z_{2G} en función de z_{1G} . Reemplazando el valor de z_{2G} en la expresión de z_{kG} da

$$z_{kG} = \frac{p(\mathbf{x}_{3t_{kG}})}{p(\mathbf{x}_{1t_{kG}})}z_{1G}. \quad (3.7)$$

Nótese que sin importar el nivel k , el valor de z_{2G} siempre es el mismo. En el caso en que G tiene sólo dos niveles, entonces las alternativas 1 y 3 son iguales. La restricción (3.4) queda de la

siguiente forma:

$$-p(\mathbf{x}_{1t_kG})z_{2G} + (p_{13}(\mathbf{x}_{1t_kG}) - p(\mathbf{x}_{1t_kG}))z_{1G} = 0 \implies z_{2G} = \frac{\frac{1}{2} - p(\mathbf{x}_{1t_kG})}{p(\mathbf{x}_{1t_kG})}z_{1G},$$

en donde $p_{13}(\mathbf{x}_{1t_kG}) = \frac{1}{2}$ pues las alternativas son iguales. Por otro lado, la restricción (3.5) queda

$$(1 - p(\mathbf{x}_{2t_kG}))z_{2G} - p(\mathbf{x}_{1t_kG})z_{1G} = 0 \implies z_{2G} = \frac{p(\mathbf{x}_{2t_kG})}{1 - p(\mathbf{x}_{2t_kG})}z_{1G}. \quad (3.8)$$

Estas dos expresiones para z_{2G} son la misma. En efecto,

$$\frac{\frac{1}{2} - p(\mathbf{x}_{1t_kG})}{p(\mathbf{x}_{1t_kG})} = \frac{\frac{1}{2} - \frac{\exp(\hat{u}_{1G})}{\exp(\hat{u}_{1G}) + \exp(\hat{u}_{2G})} \frac{1}{2}}{\frac{\exp(\hat{u}_{1G})}{\exp(\hat{u}_{1G}) + \exp(\hat{u}_{2G})} \frac{1}{2}} = \frac{\exp(\hat{u}_{2G})}{\exp(\hat{u}_{1G})} = \frac{\frac{\exp(\hat{u}_{2G})}{\exp(\hat{u}_{1G}) + \exp(\hat{u}_{2G})}}{\frac{\exp(\hat{u}_{1G})}{\exp(\hat{u}_{1G}) + \exp(\hat{u}_{2G})}} = \frac{p(\mathbf{x}_{2t_kG})}{1 - p(\mathbf{x}_{2t_kG})}.$$

Nótese que, como las probabilidades están en el intervalo $(0, 1)$, todos los cálculos anteriores son válidos. Así, todo z_{kG} depende linealmente del primer nivel de atributo z_{1G} .

Tomando ahora una pregunta t_G de las siguientes T_2 preguntas, se demuestran que dadas las condiciones sobre \hat{u} el vector \mathbf{z} queda únicamente determinado.

Cuando $n_H = 2$ y $n_G = 2$, entonces las ecuaciones (3.4) y (3.5) quedan

$$\begin{aligned} 0 &= -p(\mathbf{x}_{1t_G})(z_{2H} + z_{2G}) + (p_{13}(\mathbf{x}_{1t_G}) - p(\mathbf{x}_{1t_G}))(z_{1H} + z_{1G}) \\ 0 &= (1 - p(\mathbf{x}_{2t_G}))(z_{2H} + z_{2G}) - p(\mathbf{x}_{2t_G})(z_{1H} + z_{1G}). \end{aligned}$$

Reemplazando z_{2H} y z_{2G} según (3.8), la primera igualdad queda

$$z_{1G} \left[1 - \frac{p(\mathbf{x}_{1t_G})}{\frac{1}{2} - p(\mathbf{x}_{1t_G})} \frac{p(\mathbf{x}_{2t_kG})}{1 - p(\mathbf{x}_{2t_kG})} \right] = \frac{p(\mathbf{x}_{1t_G})}{\frac{1}{2} - p(\mathbf{x}_{1t_G})} \frac{p(\mathbf{x}_{2t_kH})}{1 - p(\mathbf{x}_{2t_kH})} - 1,$$

y la segunda igualdad queda

$$z_{1G} \left[1 - \frac{1 - p(\mathbf{x}_{2t_G})}{p(\mathbf{x}_{2t_G})} \frac{p(\mathbf{x}_{2t_kG})}{1 - p(\mathbf{x}_{2t_kG})} \right] = \frac{1 - p(\mathbf{x}_{2t_G})}{p(\mathbf{x}_{2t_G})} \frac{p(\mathbf{x}_{2t_kH})}{1 - p(\mathbf{x}_{2t_kH})} - 1.$$

Imponiendo que los coeficientes tengan el mismo signo da la condición $\hat{u}_{2G} - \hat{u}_{1G} \neq \hat{u}_{2H}$ para que \mathbf{z} quede únicamente identificado. En particular, si los coeficientes tienen signo negativo, se llega a la condición $\hat{u}_{2G} - \hat{u}_{1G} < \hat{u}_{2H}$.

Cuando $n_H = 2$ y $n_G \geq 3$, entonces las ecuaciones (3.4) y (3.5) quedan

$$\begin{aligned} 0 &= (1 - p(\mathbf{x}_{1t_G}))z_{1G} - p(\mathbf{x}_{1t_G})(z_{2H} + z_{2G} + z_{3G}) + (p_{13}(\mathbf{x}_{1t_G}) - p(\mathbf{x}_{1t_G}))z_{1H} \\ 0 &= (1 - p(\mathbf{x}_{2t_G}))(z_{2H} + z_{2G}) - p(\mathbf{x}_{2t_G})(z_{1G} + z_{3G}) - p(\mathbf{x}_{2t_G})z_{1H}. \end{aligned}$$

Reemplazando z_{2H} según (3.8) y z_{2G} y z_{3G} según (3.6) y (3.7) la primera igualdad queda

$$z_{1G} \left[1 - \frac{p(\mathbf{x}_{1t_G})}{p(\mathbf{x}_{1t_kG})} \right] = \frac{p(\mathbf{x}_{1t_G})}{1 - p(\mathbf{x}_{2t_kH})} - p_{13}(\mathbf{x}_{1t_G}),$$

y la segunda igualdad queda

$$z_{1G} \left[1 + \frac{p(\mathbf{x}_{3t_kG})}{p(\mathbf{x}_{1t_kG})} - \frac{1 - p(\mathbf{x}_{2t_G})}{p(\mathbf{x}_{2t_G})} \frac{p(\mathbf{x}_{2t_kG})}{p(\mathbf{x}_{1t_kG})} \right] = \frac{1 - p(\mathbf{x}_{2t_G})}{p(\mathbf{x}_{2t_G})} \frac{p(\mathbf{x}_{2t_kH})}{1 - p(\mathbf{x}_{2t_kG})} - 1.$$

Imponiendo que los coeficientes tengan el mismo signo da la condición $e^{\hat{u}_{1G} + \hat{u}_{2H}} + e^{\hat{u}_{2H} + \hat{u}_{3G}} \leq e^{\hat{u}_{2G}}$ para que \mathbf{z} quede únicamente identificado. En particular, si se impone que $\hat{u}_{2G} - \hat{u}_{1G} < \hat{u}_{2H}$ entonces resulta que $e^{\hat{u}_{2H} + \hat{u}_{3G}} > 0$, mientras que si $\hat{u}_{2G} - \hat{u}_{1G} > \hat{u}_{2H}$ entonces $e^{\hat{u}_{2H} + \hat{u}_{3G}} < 0$, lo que es una contradicción.

Cuando $n_H, n_G \geq 3$, entonces las ecuaciones (3.4) y (3.5) quedan

$$\begin{aligned} 0 &= (1 - p(\mathbf{x}_{1t_G}))(z_{1H} + z_{1G}) - p(\mathbf{x}_{1t_G})(z_{2H} + z_{2G} + z_{3H} + z_{3G}) \\ 0 &= (1 - p(\mathbf{x}_{2t_G}))(z_{2H} + z_{2G}) - p(\mathbf{x}_{2t_G})(z_{1H} + z_{1G} + z_{3H} + z_{3G}). \end{aligned}$$

Reemplazando z_{2H}, z_{3H}, z_{2G} y z_{3G} según (3.6) y (3.7) la primera igualdad queda

$$z_{1G} \left[1 - \frac{p(\mathbf{x}_{1t_G})}{p(\mathbf{x}_{1t_kG})} \right] = \frac{p(\mathbf{x}_{1t_G})}{p(\mathbf{x}_{1t_kH})} - 1,$$

y la segunda igualdad queda

$$z_{1G} \left[\frac{p(\mathbf{x}_{2t_kG}) - p(\mathbf{x}_{2t_G})}{p(\mathbf{x}_{1t_kG})} \right] = \frac{p(\mathbf{x}_{2t_G}) - p(\mathbf{x}_{2t_kH})}{p(\mathbf{x}_{1t_kH})}.$$

La segunda expresión anterior de z_{1G} arroja la condición $\hat{u}_{2G} - \hat{u}_{1G} \neq \hat{u}_{2H}$ para que quede únicamente determinado, condición que ya estaba impuesta. Por otro lado, la primera expresión anterior para z_{1G} arroja la siguiente condición para que quede únicamente determinado:

$$e^{\hat{u}_{1G} + \hat{u}_{2H}} - e^{\hat{u}_{2G}} \neq e^{\hat{u}_{3G}} - e^{\hat{u}_{1G} + \hat{u}_{3H}}.$$

Como esta inequidad puede ser una desigualdad estricta en cualquier sentido, basta que el coeficiente de la derecha no sea igual a cero para que quede únicamente determinado. Para que esto suceda se impone la condición $\hat{u}_{3G} - \hat{u}_{1G} \neq \hat{u}_{3H}$. \square

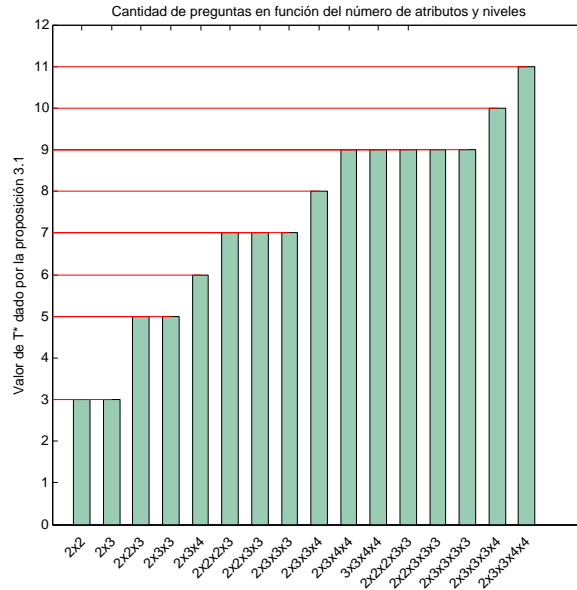


Figura 3.1: Cantidad de preguntas según número de atributos y niveles para el caso de tres alternativas.

Así como en la proposición 2.1, las restricciones impuestas sobre el vector \hat{u} acotan los posibles valores de \hat{u} en un comienzo, pero aseguran que exista un diseño con un número de preguntas razonable en la práctica, y de ahí que se imponga la condición para este estudio. Esto permite fijar un valor para T con el cual se asegura la existencia de un diseño que identifica al problema, y para efectos de la resolución, se fija $T = T^*$. La figura 3.1 muestra un gráfico con el valor de T^* en función del número de atributos.

3.5. Resultados y análisis

Se estudian en esta sección los mismos análisis que en la sección anterior: se contrastan los diseños arrojados por el algoritmo propuesto *versus* diseños compensatorios clásicos, además de propiedades inherentes del diseño. La estrategia a seguir es idéntica a la de la sección anterior: para cada cuestionario EBA se simulan respuestas con el vector \hat{u} dado en la sección 3.4, y con dichas respuestas se estiman los parámetros con la técnica de máxima verosimilitud; se calculan métricas de optimalidad para comparar diseños EBA contra diseños compensatorios clásicos; y se analizan características descriptivas.

3.5.1. Tiempo de ejecución

La tabla 3.1 muestra, para cada instancia, el número de variables enteras, el tiempo promedio de resolución de las diez veces que se resuelve el problema, y el tiempo máximo de esas diez veces. Se obtienen los mismos resultados que en el caso anterior: en promedio el algoritmo corre bastante rápido y, como es de esperar, los tiempos de resolución son más altos cuando el número de variables enteras aumenta, por lo que en un principio el problema no es escalable. Los tiempos de resolución en estas instancias particulares son aceptables, pero ocurre en contadas ocasiones que el punto de partida no permite que la resolución sea rápida, llegando al límite de los 1200 segundos.

Tabla 3.1: Número de variables y tiempo de resolución para distintas instancias.

Instancia	Número de variables enteras	Tiempo promedio de resolución [s]	Máximo tiempo de ejecución [s]
2^2	36	3,3720	7,1866
2×3	60	5,3745	17,6733
$2^2 \times 3$	126	27,2091	78,4144
2×3^2	168	17,7390	34,8194
$2 \times 3 \times 4$	216	80,7777	303,1751
$2^3 \times 3$	216	186,9792	650,3406
$2^2 \times 3^2$	270	53,8891 (168,5822) ¹	1200

1. El primer valor señala el promedio entre las instancias que no llegaron al límite de 1200 segundos; el valor entre paréntesis señala el promedio de las diez instancias.

3.5.2. Tamaño del diseño

Para el caso de tres alternativas, la tabla 3.2 resume los valores de T^* , además de mostrar el número de preguntas del diseño más pequeño, del más grande y un valor promedio entre los diez obtenidos.

Tabla 3.2: Comparación entre el número de preguntas para distintas instancias.

Instancia	T^* (prop. 3.1)	Mínimo T	Máximo T	Promedio de T
2^2	3	1	3	2,1
2×3	3	2	3	2,7
$2^2 \times 3$	5	2	5	3,7
2×3^2	5	4	5	4,7
$2 \times 3 \times 4$	6	5	6	5,3
$2^3 \times 3$	7	4	7	6,3
$2^2 \times 3^2$	7	5	7	6,2

En este caso se puede apreciar que la cota T^* fue alcanzada en todos los casos. Teniendo en cuenta que el valor de T^* en el caso de tres alternativas es menor al valor dado en el caso de dos alternativas, esto es un indicio para afirmar que a medida que hay más alternativas para mostrar el número de preguntas disminuye. Además, en general el promedio de las preguntas es más cercano al valor de T^* , lo que parece indicar que T^* sí es un número razonable de preguntas a realizar. Existen valores contraintuitivos, como en el caso 2^2 donde se obtuvo un diseño con una pregunta, o en el caso $2^2 \times 3$, donde se obtuvo uno con dos preguntas. Se debe comprobar con las métricas de D -eficiencia si estos diseños pueden en efecto ser usados para estimar los parámetros.

3.5.3. D -eficiencia y características descriptivas

Este análisis es similar al realizado en el capítulo anterior: se considera el cuestionario \mathbf{x} y el vector \mathbf{z} entregados por la resolución del problema (P), se maximiza la log-verosimilitud del modelo con \mathbf{z} como punto de partida, obteniendo un estimador de los pesos del modelo $\hat{\alpha}_{\mathbf{x}}$, y se calcula la matriz de información de Fisher del modelo y su determinante normalizado. Sin embargo, la matriz de información de Fisher tiene una expresión distinta, pues las probabilidades de elección cambian. En el apéndice C, sección C.2 se provee la forma de dicha matriz.

A modo de ejemplo, la tabla 3.3 muestra, para los diez cuestionarios obtenidos con el algoritmo propuesto en un contexto fijo, su número de preguntas y su D^{EBA} -error.

Se pueden apreciar resultados diversos: por un lado hay diseños con error infinito, que se traduce en que la matriz de información asociada tiene determinante igual a cero; hay diseños con errores grandes (en este caso, del orden de 10^3 y 10^5); y diseños con errores muy pequeños, similares al caso anterior.

Tabla 3.3: D^{EBA} -error para el caso $2 \times 3 \times 3$ versus número de preguntas.

Número de cuestionario	Número de preguntas	D^{EBA} -error
1	5	$1,0053 \times 10^5$
2	5	$4,1571 \times 10^5$
3	4	∞
4	5	$1,2524 \times 10^3$
5	5	$2,1941 \times 10^3$
6	4	$3,2557 \times 10^{-18}$
7	5	$9,1297 \times 10^{-19}$
8	5	∞
9	4	$5,0807 \times 10^{-17}$
10	5	$1,1090 \times 10^5$

La tabla 3.4 resume, para cada contexto de niveles y atributos, el D^{EBA} -error más pequeño encontrado entre los diez cuestionarios, su número de preguntas y el D^{EBA} -error asociado al cuestionario dado por la proposición 3.1. Así como en el caso anterior, no se puede obtener una conclusión decidora con respecto a qué tanto mejor es el cuestionario dado en la proposición. En algunos contextos de atributos este cuestionario resultó ser mejor (en el sentido de la D -eficiencia) que el mejor cuestionario arrojado por la resolución de (P), pero en otros casos sucede lo contrario.

Tabla 3.4: Datos de los mejores cuestionarios EBA por contexto.

Contexto	Número de preguntas	Mejor D^{EBA} -error	Número de preguntas (prop 2.1)	D^{EBA} -error (prop 2.1)
2^2	2	$1,4858 \times 10^{-6}$	3	∞
2×3	3	$7,4146 \times 10^{-4}$	3	0,0016
$2^2 \times 3$	4	$2,1887 \times 10^{-4}$	5	$3,1113 \times 10^{-4}$
2×3^2	5	$9,1297 \times 10^{-19}$	5	$4,2369 \times 10^{-17}$
$2 \times 3 \times 4$	5	$5,7912 \times 10^{-20}$	6	$2,6747 \times 10^{-21}$
$2^3 \times 3$	6	$2,1065 \times 10^{-11}$	7	$4,2889 \times 10^{-13}$
$2^2 \times 3^2$	7	466,0759	7	771,3994

Con respecto a la pérdida de eficiencia, se usan las definiciones dadas en la sección 2.5.3, análogas para este caso, considerando la matriz de información correspondiente al caso de tres alternativas. Se analizan a continuación estas métricas para concluir qué tan eficiente es usar un cuestionario cuando el modelo subyacente es otro.

Para cada contexto se reportan las medidas de pérdida de eficiencia en la tabla 3.5. Cuando el modelo subyacente es EBA (tabla 3.5a), se concluye que en casi todos los casos los diseños compensatorios clásicos presentan altas pérdidas de eficiencia cuando la métrica base es el D^{EBA} -error. El único contexto anómalo es 2×3^2 , en donde el diseño compensatorio resultó ser mejor que el diseño EBA. Por otro lado, las pérdidas de eficiencia basadas en el RMSE dependen bastante del contexto.

Tabla 3.5: Pérdida de eficiencia según distintas métricas.

(a) El modelo subyacente es EBA			(b) El modelo subyacente es compensatorio		
Contexto	Porcentaje de pérdida de eficiencia basada en		Contexto	Porcentaje de pérdida de eficiencia basada en	
	$D^{\text{EBA}}\text{-error}$	RMSE		$D^{\text{Comp}}\text{-error}$	RMSE
2^2	63,4592	74,4387	2^2	32,7287	$0(-13,2697)^1$
2×3	99,8409	0,0374	2×3	16,2176	$0(-52,5368)^1$
$2^2 \times 3$	99,9904	$0(-85,9108)^1$	$2^2 \times 3$	44,7900	22,0895
2×3^2	$0(-15,5911)^1$	$3,3056 \times 10^{-4}$	2×3^2	59,0269	34,1929
$2 \times 3 \times 4$	94,0547	$0(-0,0187)^1$	$2 \times 3 \times 4$	51,9891	79,7124
$2^3 \times 3$	100	$0(-17,3063)^1$	$2^3 \times 3$	19,7906	28,1589
$2^2 \times 3^2$	99,9743	$0(-16,3761)^1$	$2^2 \times 3^2$	39,7017	53,3201

1. Valor entre paréntesis: pérdida de eficiencia efectiva según la fórmula. Como el número es negativo, la pérdida de eficiencia es 0.

1. Valor entre paréntesis: pérdida de eficiencia efectiva según la fórmula. Como el número es negativo, la pérdida de eficiencia es 0.

Si el modelo subyacente es compensatorio (tabla 3.5b), se concluye que los diseños EBA tienen una menor pérdida que las contrapartes compensatorias si la métrica base es el $D^{\text{Comp}}\text{-error}$. Así como en el análisis anterior, sucede que las pérdidas de eficiencia basadas en el RMSE dependen del contexto.

Los resultados finales de esta sección también indican conclusiones algo contradictorias. Una vez más ocurre que, dependiendo de la métrica usada, los diseños EBA resultan ser mejores o peores que los compensatorios a la hora de estimar los parámetros del modelo. La métrica usual (D -eficiencia) favorece a los diseños EBA a la hora de medir pérdidas de eficiencia. Sin embargo, en este caso la métrica que refleja distancia entre parámetros reales y estimados se comporta de manera errática.

3.5.4. Características descriptivas

Se analizan ahora las mismas características descriptivas mencionadas en la sección 2.5.4 para el caso de tres alternativas: balance de niveles, ortogonalidad y traslape de niveles.

Así como en el caso de dos alternativas, las figuras 3.2, 3.3 y 3.4 confirman que los diseños EBA en promedio presentan menor balance de niveles, y un mayor traslape de niveles que sus contrapartes compensatorias. Por otro lado, el comportamiento de la ortogonalidad de los diseños es errático: en algunos casos los diseños EBA resultaron ser «más ortogonales» que los diseños compensatorios, pero en otros casos no.

El gráfico 3.4 permite confirmar lo planteado en el análisis análogo hecho en el caso de dos alternativas: si existe un nivel que se traslapa entre las tres alternativas, entonces éste no participa de las probabilidades de elección, por lo que son menos los parámetros que actúan en esa pregunta. Por otro lado, si un nivel aparece en dos de tres alternativas, entonces aporta al traslape, pero no se dificulta la resolución del problema de optimización: en efecto, ya sea si el nivel aparece en sólo una alternativa o en dos, la variable asociada aún está presente en las restricciones, y como

no se impone que haya o no traslape, no es de extrañar que la característica esté presente en los diseños EBA.

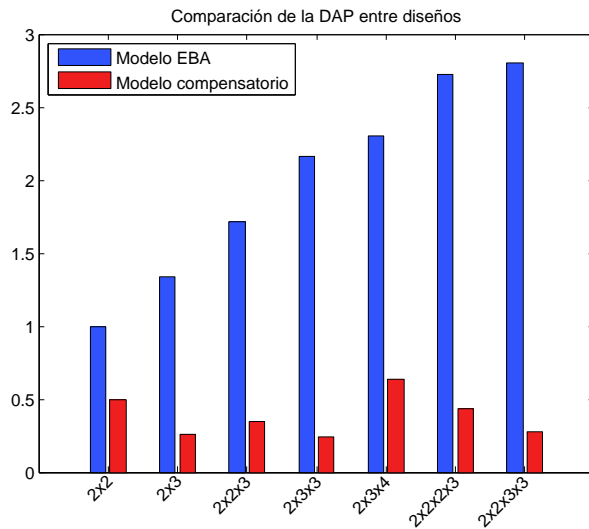


Figura 3.2: Comparación de la DAP.

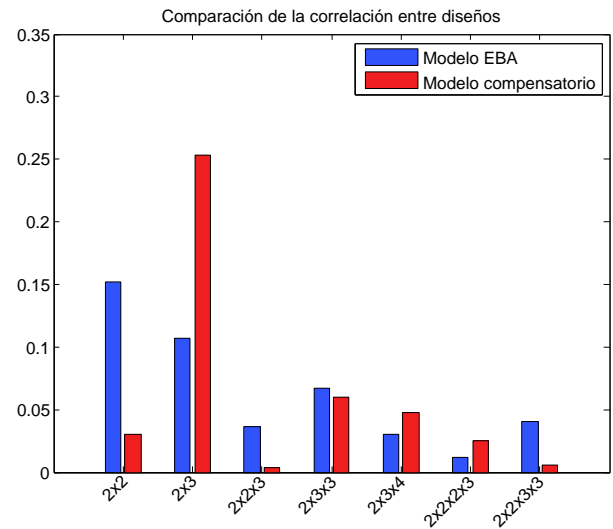


Figura 3.3: Comparación de la correlación.

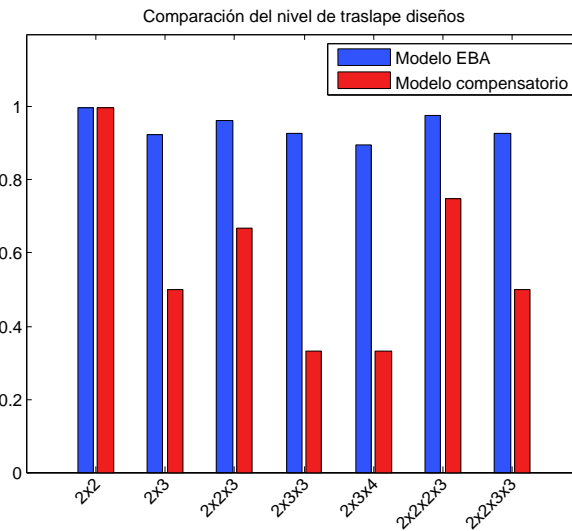


Figura 3.4: Comparación del nivel de traslape.

El balance de niveles también sigue la misma lógica que el caso de dos alternativas: puede ocurrir que un nivel k aparezca muchas veces, y todos los otros niveles queden en función de k , así como ocurre en el diseño construido en la proposición 3.1.

Conclusión

El principal resultado teórico de este trabajo es el planteamiento de un problema de optimización MIP binivel que, vía su resolución heurística, entrega diseños para un contexto EBA que, según la métrica de la D -eficiencia, es mejor que los diseños compensatorios clásicos para identificar los parámetros del modelo.

Con respecto al problema de optimización, se aprovecha sobremanera la estructura de las probabilidades de elección y que gracias a éstas se puede plantear un sistema lineal (si bien se suponen conocidas dichas probabilidades). La resolución del problema usa técnicas clásicas de generación de columnas y restricciones, además de trabajar con variables mixtas, para lo cual existen herramientas especializadas que permiten su resolución de manera rápida. Así, se logran construir cuestionarios acordes al modelo EBA, uno de los objetivos finales de la tesis.

Los análisis descriptivos y métricas asociadas permiten concluir que, desde el punto de vista de la D -eficiencia, en general los cuestionarios obtenidos vía el problema de optimización son mejores que los compensatorios clásicos a la hora de recuperar los parámetros del modelo. Sin embargo, cuando la métrica cambia a una de distancia entre el parámetro real y el estimado, los diseños compensatorios funcionan mejor. La ortogonalidad, traslape y balance de niveles, deseables en los diseños compensatorios, no están presentes en los diseños EBA. Dichas características no se buscaban, pero se confirma que no es necesario hacerlo por cómo funciona EBA.

Los resultados con respecto a los tiempos de resolución permiten concluir que el método implementado no es aplicable a instancias más grandes sin alguna modificación a la implementación del MIP. Los tiempos documentados son aceptables, pero si el objetivo es resolver esto de forma más rápida en instancias más grandes (como en el ejemplo de las cámaras fotográficas, donde la cantidad de aspectos es 4^5) y comunes en aplicaciones se debe modificar el método de resolución o la descripción del problema *per se*.

Una primera extensión obvia a desarrollar es el planteamiento para más alternativas, y la idea es la misma que se aplica para el caso de tres alternativas: conociendo las probabilidades de elección, se puede plantear un sistema lineal en las variables correspondientes y, así, plantear el subproblema que permite identificar.

Un segundo resultado a generalizar concierne a proveer una cota en el número de preguntas para el caso de n alternativas, así como se hizo en las proposiciones 2.1 y 3.1, siempre considerando que el objetivo de esto es facilitar la resolución de (P) al imponer un número de restricciones.

Otro resultado a generalizar es la eliminación de las condiciones impuestas por las proposiciones 2.1 y 3.1 sobre los valores que pueden tomar los parámetros del problema. Se conjetura que, al eliminar las condiciones, aún se pueda obtener un cuestionario que permita identificar el modelo y, en el peor de los casos, el número de preguntas sea más grande que el provisto en las proposiciones.

Conjetura

Si el vector $\hat{\mathbf{u}}$ no cumple con las condiciones impuestas en las proposiciones (2.1) y (3.1), entonces existe un cuestionario con $T > T^*$ preguntas que permite identificar el modelo.

Otro ambiente en donde es provechoso seguir profundizando es la técnica usada al resolver el problema de optimización. Por un lado, la heurística planteada para resolver el problema MIP binivel puede no ser la adecuada y, por ende, refinada para acelerar la resolución. Si sólo se considera aplicar las restricciones del tipo (2.13), el algoritmo se vuelve uno de fuerza bruta que revise todos los diseños posibles.

Por último, existen otros métodos de plantear problemas de optimización para obtener diseños que pueden ser adaptados al caso EBA. Como se menciona en la sección 1.5, los trabajos de Sagnol y Harman [SH15] y Duarte y Wong [DW15] usan técnicas de programación cónica de segundo orden y programación semidefinida positiva para construir diseños D -óptimos. Si bien ambos trabajos se basan en modelos de regresión, es posible que las técnicas empleadas puedan ser adaptadas para tener en cuenta el modelo EBA.

Bibliografía

- [AdT92] Simon Anderson, André de Palma y Jacques-François Thisse: *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge, 1992.
- [AHJS97] Charles Audet, Pierre Hansen, Brigitte Jaumard y Gilles Savard: *Links Between Linear Bilevel and Mixed 0–1 Programming Problems*. Journal of Optimization Theory and Applications, 93(2):273–300, 1997.
- [Bai08] Rosemary Bailey: *Design of Comparative Experiments*, volumen 25. Cambridge University Press, New York, 2008.
- [BLA94] David Bunch, Jordan Louviere y Don Anderson: *A Comparison of Experimental Design Strategies for Multinomial Logit Models: The Case of Generic Attributes*. Graduate School of Management, Universidad de California en Davis, 1994.
- [BM73] Jerome Bracken y James T. McGill: *Mathematical Programs with Optimization Problems in the Constraints*. Operations Research, 21(1):37–44, 1973.
- [BT97] Dimitris Bertsimas y John Tsitsiklis: *Introduction to Linear Optimization*, volumen 6. Athena Scientific, Nashua, 1997.
- [CAV13] Marjolein Crabbe, Deniz Akinç y Martina Vandebroek: *Fast algorithms to generate individualized designs for the mixed logit choice model*. Transportation Research Part B: Methodological, 60:1–15, 2013.
- [CN77] Wilfred Candler y Roger Norton: *Multi-level Programming and Development Policy*. The World Bank, Washington, D.C., 1977.
- [Deb60] Gerard Debreu: *Review of Individual Choice Behavior by R. Luce*. American Economic Review, 50:186–188, 1960.
- [DMSB15] Angela Dean, Max Morris, John Stufken y Derek Bingham (editores): *Handbook of Design and Analysis of Experiments*, volumen 7. CRC Press, Boca Raton, 2015.
- [DW15] Belmiro P.M. Duarte y Weng Kee Wong: *Finding Bayesian Optimal Designs for Non-linear Models: A Semidefinite Programming-Based Approach*. International Statistical Review, 83(2):239–262, 2015.
- [Fra95] Antonio Frangioni: *On a new class of bilevel programming problems and its use for reformulating mixed integer problems*. European Journal of Operational Research, 82(3):615–646, 1995.
- [HZ96] Joel Huber y Klaus Zwerina: *The Importance of Utility Balance in Efficient Choice Designs*. Journal of Marketing Research, 33(3):307–317, 1996.
- [IBM12] IBM® Corp., Armonk, NY: *IBM SPSS Statistics 21 Core System User's Guide*, 1ª edición, 2012.

- [KJM14] Rajeev Kohli, Kamel Jedidi y Ricardo Montoya: *Elimination By Aspects as a Logit Model*. Departamento de Ingeniería Industrial, Universidad de Chile. Por publicar, 2014.
- [KTG94] Warren Kuhfeld, Randall Tobias y Mark Garratt: *Efficient Experimental Design with Marketing Research Applications*. *Journal of Marketing Research*, 31(4):545–557, 1994.
- [Kuh10] Warren Kuhfeld: *Marketing Research Methods in SAS*. Nota técnica, SAS Institute Inc., octubre 2010.
- [LA11] Qing Liu y Neeraj Arora: *Efficient Choice Designs for a Consider-Then-Choose Model*. *Marketing Science*, 30(2):321–338, 2011.
- [LS65] Robert Luce y Patrick Suppes: *Preference, Utility and Subjective Probability*. En Robert Luce, Robert Bush y Eugene Galanter (editores): *Handbook of Mathematical Psychology*, páginas 249–410. Wiley, New York, 1965.
- [Luc59] Robert Luce: *Individual Choice Behavior*. Wiley, New York, 1959.
- [Man77] Charles Manski: *The Structure of Random Utility Models*. *Theory and Decision*, 8(3): 229–254, 1977.
- [Mar10] François Margot: *Symmetry in Integer Linear Programming*. En Michael Jünger, Thomas Liebling, Denis Naddef, George Nemhauser, William Pulleyblank, Gerhard Reinelt, Giovanni Rinaldi y Laurence Wosley (editores): *50 Years of Integer Programming 1958-2008*, páginas 647–686. Springer Berlin Heidelberg, Dordrecht, London, New York, 2010.
- [McF73] Daniel McFadden: *Conditional Logit Analysis of Qualitative Choice Behavior*. En Paul Zarembka (editor): *Frontiers in Econometrics*, páginas 105–142. Academic Press, New York, 1973.
- [Qua56] Richard Quandt: *A Probabilistic Theory of Consumer Behavior*. *Quarterly Journal of Economics*, 70(4):507–536, 1956.
- [SH15] Guillaume Sagnol y Radoslav Harman: *Computing exact D-optimal designs by mixed integer second-order cone programming*. *The Annals of Statistics*, 43(5):2198–2224, 2015.
- [SW01] Zsolt Sándor y Michel Wedel: *Designing conjoint choice experiments using managers' prior beliefs*. *Journal of Marketing Research*, 38(4):430–444, 2001.
- [SW02] Zsolt Sándor y Michel Wedel: *Profile Construction in Experimental Choice Designs for Mixed Logit Models*. *Marketing Science*, 21(4):455–475, 2002.
- [TH07] Olivier Toubia y John Hauser: *On Managerially Efficient Experimental Designs*. *Marketing Science*, 26(6):851–858, 2007.
- [Thu27a] Louis Thurstone: *A Law of Comparative Judgement*. *Psychological Review*, 34(4): 273–286, 1927.
- [Thu27b] Louis Thurstone: *Psychological Analysis*. *American Journal of Psychology*, 38(3): 368–389, 1927.
- [Thu45] Louis Thurstone: *The Prediction of Choice*. *Psychometrika*, 10(4):237–253, 1945.

- [Tra09] Kenneth Train: *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, 2^a edición, 2009.
- [Tve72a] Amos Tversky: *Choice by Elimination*. Journal of Mathematical Psychology, 9(4): 341–367, 1972.
- [Tve72b] Amos Tversky: *Elimination by Aspects: A Theory of Choice*. Psychological Review, 79(4):281–299, 1972.
- [ZA14] Bo Zeng y Yu An: *Solving Bilevel Mixed Integer Program by Reformulations and Decomposition*. Optimization On-Line, 2014.

Apéndice A

Propiedades de la distribución Gumbel

Como se menciona en la subsección 1.1.2, la utilidad de un individuo posee una componente de error ε que representa la incertidumbre en la utilidad que se le da a un aspecto o característica, y captura el hecho de que la importancia de tal característica depende del contexto de la elección o el marco mental del individuo. Por ejemplo, una persona puede a veces preferir chocolate dulce y a veces chocolate amargo, aún cuando sea muy amante del chocolate dulce. Hay veces, incluso, en que este error refleja dudas frente a elecciones que son hechas sólo una vez.

Las distribuciones de valor extremo se usan para modelar las situaciones descritas en el párrafo anterior. La distribución Gumbel o doble exponencial es una de las más importantes de esta familia, por su simplicidad y por las propiedades que posee. Algunas propiedades importantes son mostradas a continuación. Considérese una distribución Gumbel con parámetro de locación $\eta \in \mathbb{R}$ y parámetro de escala $\mu > 0$,

$$F(x) = \exp \left[- \exp \left(- \frac{x - \eta}{\mu} \right) \right]. \quad (\text{A.1})$$

Se puede demostrar que la esperanza de la doble exponencial es $\mu\gamma + \eta$, con γ la constante de Euler-Mascheroni. Así, para que la media sea igual a cero debe cumplirse que $\eta = -\mu\gamma$ (como ocurre en (1.9)). Por otro lado, la varianza es $\sigma^2 = \frac{\mu^2\pi^2}{6}$.

Si ε_1 y ε_2 son dos variables aleatorias distribuidas acordes a (A.1), entonces $\mathbb{P}(\varepsilon_2 - \varepsilon_1 \leq x) = (1 + \exp(-x/\mu))^{-1}$, que es la función de distribución de una variable logística de media cero y varianza $\frac{\mu^2\pi^2}{3}$.

Si $\varepsilon_1, \dots, \varepsilon_n$ son variables aleatorias i.i.d. con parámetro de escala μ y se definen las variables $\tilde{U}_1, \dots, \tilde{U}_n$ como en la subsección (1.1.2), se puede demostrar que

$$\mathbb{E} \left(\max_{i=1, \dots, n} \tilde{U}_i \right) = \mu \ln \left[\sum_{i=1}^n \exp \left(\frac{u_i}{\mu} \right) \right].$$

Además, si los parámetros de locación son η_i para cada i , entonces

$$\mathbb{P} \left(\max_{i=1, \dots, n} \varepsilon_i \leq x \right) = \exp \left[- \exp \left(- \frac{x - \bar{\eta}}{\mu} \right) \right],$$

donde $\exp(\bar{\eta}/\mu) = \sum_{i=1}^n \exp(\eta_i/\mu)$. Así, el máximo de variables aleatorias Gumbel es Gumbel.

Apéndice B

Algoritmo de generación de columnas para el problema BiMIP y adaptación

El siguiente apéndice es un resumen del artículo de Zeng y An [ZA14].

B.1. El algoritmo general

Un problema MIP binivel general tiene la forma

$$\begin{aligned} \text{BiMIP} : \Theta^* &= \text{mín } \mathbf{f}\mathbf{x} + \mathbf{g}\mathbf{y} + \mathbf{h}\mathbf{z} \\ \text{sujeto a } &\mathbf{A}\mathbf{x} \leq \mathbf{b}, \\ &(\mathbf{y}, \mathbf{z}) \in \arg \text{máx}\{\mathbf{w}\mathbf{y} + \mathbf{v}\mathbf{z} : \mathbf{P}\mathbf{y} + \mathbf{N}\mathbf{z} \leq \mathbf{R} - \mathbf{K}\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{n_c}, \mathbf{z} \in \mathbb{Z}_+^{n_d}\}, \\ &\mathbf{x} \in \mathbb{R}_+^{m_c} \times \mathbb{Z}_+^{m_d}, \end{aligned} \quad (\text{B.1})$$

en donde \mathbf{x} representa las variables de decisión de nivel superior, \mathbf{y} representa las variables de decisión continuas de nivel inferior y \mathbf{z} representa las variables de decisión discretas de nivel inferior. Es sabido que los problemas MIP biniveles son difíciles de resolver. Incluso la versión más simple del problema, en donde ambos problemas superior e inferior son lineales, es teóricamente NP-difícil [AHJS97, Fra95]. Sin embargo, Zeng y An desarrollan una heurística en forma de algoritmo para este tipo de problemas, basándose en reformular la instancia original y aplicar generación de columnas y restricciones. Se procede a resumir el resultado de dicho artículo y el algoritmo a aplicar, siguiendo su propia notación.

El primer paso es duplicar las variables y restricciones del problema de nivel inferior. La formulación de dicho problema, equivalente a la anterior, es

$$\text{BiMIP}_d : \Theta^* = \text{mín } \mathbf{f}\mathbf{x} + \mathbf{g}\mathbf{y}^0 + \mathbf{h}\mathbf{z}^0 \quad (\text{B.2})$$

$$\text{sujeto a } \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \in \mathbb{R}_+^{m_c} \times \mathbb{Z}_+^{m_d}, \quad (\text{B.3})$$

$$\mathbf{P}\mathbf{y}^0 + \mathbf{N}\mathbf{z}^0 \leq \mathbf{R} - \mathbf{K}\mathbf{x}, \mathbf{y}^0 \in \mathbb{R}_+^{n_c}, \mathbf{z}^0 \in \mathbb{Z}_+^{n_d}, \quad (\text{B.4})$$

$$\mathbf{w}\mathbf{y}^0 + \mathbf{v}\mathbf{z}^0 \geq \text{máx}\{\mathbf{w}\mathbf{y} + \mathbf{v}\mathbf{z} : \mathbf{P}\mathbf{y} + \mathbf{N}\mathbf{z} \leq \mathbf{R} - \mathbf{K}\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{n_c}, \mathbf{z} \in \mathbb{Z}_+^{n_d}\}, \quad (\text{B.5})$$

El segundo paso es separar las variables discretas y continuas del problema de nivel inferior para reestructurar el lado derecho de (B.5) como sigue:

$$\mathbf{w}\mathbf{y}^0 + \mathbf{v}\mathbf{z}^0 \geq \text{máx}_{\mathbf{z} \in \mathbb{Z}} \mathbf{v}\mathbf{z} + \text{máx}\{\mathbf{w}\mathbf{y} : \mathbf{P}\mathbf{y} \leq \mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z}, \mathbf{y} \in \mathbb{R}_+^{n_c}\}, \quad (\text{B.6})$$

donde \mathbf{Z} es la colección de todos los posibles \mathbf{z} . De buenas a primeras esto es inútil porque dificulta el problema, pero como el segundo problema de maximización en (B.6) es un problema lineal puro se pueden usar las condiciones de Karush-Kuhn-Tucker (KKT) para tener una formulación equivalente.

$$\mathbf{w}\mathbf{y}^0 + \mathbf{v}\mathbf{z}^0 \geq \max_{\mathbf{z} \in \mathbf{Z}} \mathbf{v}\mathbf{z} + \mathbf{w}\mathbf{y}$$

$$\text{sujeto a } (\mathbf{y}, \boldsymbol{\pi}) \in \{\mathbf{P}\mathbf{y} \leq \mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z}, \mathbf{P}^\top \boldsymbol{\pi} \geq \mathbf{w}^\top, \mathbf{y} \perp (\mathbf{P}^\top \boldsymbol{\pi} - \mathbf{w}^\top), \boldsymbol{\pi} \perp (\mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z} - \mathbf{P}\mathbf{y})\}.$$

Enumerando $\mathbf{Z} = \{\mathbf{z}^1, \dots, \mathbf{z}^k\}$, declarando variables $(\mathbf{y}^j, \boldsymbol{\pi}^j)$ y las condiciones de KKT relacionadas se tiene el siguiente teorema.

Teorema B.1

La formulación BiMIP_d en (B.2 – B.5) equivale a la siguiente formulación expandida, de sólo un nivel:

$$\Sigma_{\mathbf{Z}} : \text{mín } \mathbf{f}\mathbf{x} + \mathbf{g}\mathbf{y}^0 + \mathbf{h}\mathbf{z}^0 \tag{B.7}$$

$$\text{sujeto a } \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \in \mathbb{R}_+^{m_c} \times \mathbb{Z}_+^{m_d}, \tag{B.8}$$

$$\mathbf{P}\mathbf{y}^0 + \mathbf{N}\mathbf{z}^0 \leq \mathbf{R} - \mathbf{K}\mathbf{x}, \tag{B.9}$$

$$\mathbf{w}\mathbf{y}^0 + \mathbf{v}\mathbf{z}^0 \geq \mathbf{w}\mathbf{y}^j + \mathbf{v}\mathbf{z}^j, \quad 1 \leq j \leq k, \tag{B.10}$$

$$\mathbf{P}\mathbf{y}^j \leq \mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z}^j, \mathbf{P}^\top \boldsymbol{\pi}^j \geq \mathbf{w}^\top, \quad 1 \leq j \leq k, \tag{B.11}$$

$$\mathbf{y}^j \perp (\mathbf{P}^\top \boldsymbol{\pi}^j - \mathbf{w}^\top), \boldsymbol{\pi}^j \perp (\mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z}^j - \mathbf{P}\mathbf{y}^j), \quad 1 \leq j \leq k, \tag{B.12}$$

$$\mathbf{y}^j \in \mathbb{R}_+^{n_c}, \boldsymbol{\pi}^j \in \mathbb{R}_+^{n_1}, \quad 1 \leq j \leq k, \tag{B.13}$$

$$\mathbf{y}^0 \in \mathbb{R}_+^{n_c}, \mathbf{z}^0 \in \mathbb{Z}_+^{n_d}. \tag{B.14}$$

La formulación anterior es un programa matemático con restricciones de complementariedad, que puede ser fácilmente convertido en un problema MIP regular añadiendo variables binarias adicionales [AHJS97]. En específico, una restricción de complementariedad $\pi_i(\mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{P}\mathbf{y})_i = 0$ equivale a

$$\pi_i \leq L\delta_i, \quad (\mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{P}\mathbf{y})_i \leq L(1 - \delta_i), \quad \delta_i \in \{0, 1\},$$

con L una constante lo suficientemente grande. Con esta técnica, $\Sigma_{\mathbf{Z}}$ puede ser tratado y resuelto como un MIP.

Zeng y An dan además reformulaciones alternativas del BiMIP_d . Se puede usar el teorema de dualidad fuerte en (B.6) para obtener otra formulación equivalente al problema original. También mencionan los llamados modelos de interdicción de la forma

$$\Theta^* = \min_{\mathbf{x}} \max \{ \mathbf{w}\mathbf{y} + \mathbf{v}\mathbf{z} : (\mathbf{y}, \mathbf{z}) \in \mathcal{F}(\mathbf{x}) \}, \tag{B.15}$$

con $\mathcal{F}(\mathbf{x}) = \{ \mathbf{y} \in \mathbb{R}_+^{n_c}, \mathbf{z} \in \mathbb{Z}_+^{n_d} : \mathbf{P}\mathbf{y} + \mathbf{N}\mathbf{z} \leq \mathbf{R} - \mathbf{K}\mathbf{x} \}$. Los problemas planteados en los capítulos 2 y 3 caen dentro de esta clase de problemas. Para éstos se tiene el siguiente corolario.

Corolario B.2

El modelo de interdicción (B.15) es equivalente a la siguiente formulación de sólo un nivel:

$$\Sigma_{\mathbf{Z}}^I : \text{mín } \{ \eta : \eta \geq \mathbf{v}\mathbf{z}^j + \mathbf{w}\mathbf{y}^j, 1 \leq j \leq k, \text{ (B.3), (B.11) – (B.13)} \}.$$

El algoritmo del artículo usa generación de columnas y restricciones. La idea es la siguiente: considérese una solución de nivel superior \mathbf{x}^* . Al resolver el subproblema, que es el problema de nivel inferior, se deriva un óptimo $(\mathbf{y}^*, \mathbf{z}^*)$. Como $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*)$ es una solución factible, su valor $\mathbf{f}\mathbf{x}^* + \mathbf{g}\mathbf{y}^* + \mathbf{h}\mathbf{z}^*$ provee una cota superior a \mathbf{BiMIP}_d . Se actualiza el conjunto de los posibles \mathbf{z} agregando \mathbf{z}^* y se expande el problema maestro. Resolver el problema maestro da un nuevo \mathbf{x}^* y una cota inferior más fuerte para una nueva iteración.

Sean UB y LB las cotas superiores e inferiores respectivamente, l el índice de la iteración y ε la tolerancia para la optimalidad. El algoritmo se muestra a continuación.

Algoritmo 4 Generación de columnas y restricciones para MIP binivel.

- 1: $LB = -\infty, UB = \infty, l = 0$.
- 2: Resolver el siguiente problema maestro:

$$\begin{aligned} \text{MP : } \underline{\Theta}^* &= \text{mín } \mathbf{f}\mathbf{x} + \mathbf{g}\mathbf{y}^0 + \mathbf{h}\mathbf{z}^0 \\ \text{sueto a } \mathbf{A}\mathbf{x} &\leq \mathbf{b}, \mathbf{x} \in \mathbb{R}_+^{m_c} \times \mathbb{Z}_+^{m_d}, \\ \mathbf{P}\mathbf{y}^0 + \mathbf{N}\mathbf{z}^0 &\leq \mathbf{R} - \mathbf{K}\mathbf{x}, \\ \mathbf{w}\mathbf{y}^0 + \mathbf{v}\mathbf{z}^0 &\geq \mathbf{w}\mathbf{y}^j + \mathbf{v}\mathbf{z}^j, & 1 \leq j \leq l, \\ \mathbf{P}\mathbf{y}^j &\leq \mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z}^j, \mathbf{P}^\top \boldsymbol{\pi}^j \geq \mathbf{w}^\top, & 1 \leq j \leq l, \\ \mathbf{y}^j &\perp (\mathbf{P}^\top \boldsymbol{\pi}^j - \mathbf{w}^\top), \boldsymbol{\pi}^j \perp (\mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z}^j - \mathbf{P}\mathbf{y}^j), & 1 \leq j \leq l, \\ \mathbf{y}^j &\in \mathbb{R}_+^{n_c}, \boldsymbol{\pi}^j \in \mathbb{R}_+^{n_1}, & 1 \leq j \leq l, \\ \mathbf{y}^0 &\in \mathbb{R}_+^{n_c}, \mathbf{z}^0 \in \mathbb{Z}_+^{n_d}. \end{aligned}$$

- Derivar una solución óptima $(\mathbf{x}^*, \mathbf{y}^{0*}, \mathbf{z}^{0*}, \mathbf{y}^{1*}, \dots, \mathbf{y}^{l*}, \boldsymbol{\pi}^{1*}, \dots, \boldsymbol{\pi}^{l*})$ y actualizar $LB = \underline{\Theta}^*$.
- 3: **si** $UB - LB \leq \varepsilon$ **entonces devolver** UB y la solución correspondiente (incumbente).
- 4: **si no**
- 5: Resolver el siguiente problema para \mathbf{x}^* , que sirve como el primer subproblema:

$$\theta(\mathbf{x}^*) = \text{máx}\{\mathbf{w}\mathbf{y} + \mathbf{v}\mathbf{z} : \mathbf{P}\mathbf{y} + \mathbf{N}\mathbf{z} \leq \mathbf{R} - \mathbf{K}\mathbf{x}^*, \mathbf{y} \in \mathbb{R}_+^{n_c}, \mathbf{z} \in \mathbb{Z}_+^{n_d}\}.$$

Resolver luego el siguiente problema para \mathbf{x}^* , que sirve como el segundo subproblema:

$$\Theta_o(\mathbf{x}^*) = \text{mín}\{\mathbf{g}\mathbf{y} + \mathbf{h}\mathbf{z} : \mathbf{w}\mathbf{y} + \mathbf{v}\mathbf{z} \geq \theta(\mathbf{x}^*), \mathbf{P}\mathbf{y} + \mathbf{N}\mathbf{z} \leq \mathbf{R} - \mathbf{K}\mathbf{x}^*, \mathbf{y} \in \mathbb{R}_+^{n_c}, \mathbf{z} \in \mathbb{Z}_+^{n_d}\}.$$

- Derivar una solución óptima $(\mathbf{y}^*, \mathbf{z}^*)$ y actualizar $UB \leftarrow \text{mín}\{UB, \mathbf{f}\mathbf{x}^* + \Theta_o(\mathbf{x}^*)\}$.
- 6: Fijar $\mathbf{z}^{l+1} = \mathbf{z}^*$, crear variables $(\mathbf{y}^{l+1}, \boldsymbol{\pi}^{l+1})$ y añadir las siguientes restricciones a **MP**:

$$\begin{aligned} \mathbf{w}\mathbf{y}^0 + \mathbf{v}\mathbf{z}^0 &\geq \mathbf{w}\mathbf{y}^{l+1} + \mathbf{v}\mathbf{z}^{l+1}, \mathbf{P}\mathbf{y}^{l+1} \leq \mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z}^{l+1}, \mathbf{P}^\top \boldsymbol{\pi}^{l+1} \geq \mathbf{w}^\top, \\ \mathbf{y}^{l+1} &\perp (\mathbf{P}^\top \boldsymbol{\pi}^{l+1} - \mathbf{w}^\top), \boldsymbol{\pi}^{l+1} \perp (\mathbf{R} - \mathbf{K}\mathbf{x} - \mathbf{N}\mathbf{z}^{l+1} - \mathbf{P}\mathbf{y}^{l+1}), \\ \mathbf{y}^{l+1} &\in \mathbb{R}_+^{n_c}, \boldsymbol{\pi}^{l+1} \in \mathbb{R}_+^{n_1}. \end{aligned}$$

Actualizar $l \leftarrow l + 1$ e ir al paso 2.

- 7: **fin si**
-

El algoritmo converge en finitas iteraciones, como muestra el siguiente resultado.

Proposición B.3

Sea $\varepsilon = 0$ y supóngase que $|\mathbf{Z}| < \infty$. El algoritmo presentado de generación de columnas y restricciones converge al valor óptimo de \mathbf{BiMIP}_d en $O(|\mathbf{Z}|)$ iteraciones. Más aún, si una solución finita óptima existe, entonces el algoritmo converge a una solución óptima en finitas iteraciones, sin importar si $|\mathbf{Z}| = \infty$.

Por último, si bien este algoritmo está hecho para un problema \mathbf{BiMIP} general, el caso particular del presente trabajo es un modelo de interdicción (B.15), donde se «minimiza un máximo». Esta estructura permite simplificar el algoritmo, ya que para calcular el problema maestro se usa $\Sigma_{\mathbf{Z}}^I$, no se declaran las variables $\mathbf{y}^0, \mathbf{z}^0$ y se omite el cálculo del segundo subproblema del algoritmo, derivando la solución óptima $(\mathbf{y}^*, \mathbf{z}^*)$ del primer subproblema y actualizando $UB \leftarrow \min\{UB, \theta(\mathbf{x}^*)\}$.

B.2. Adaptación al problema de la tesis

El problema (P), una vez hechas todas las modificaciones de las secciones 2.4 y 3.4, no es exactamente un problema BIMIP de la forma (B.1): la matriz \mathbf{P} en (B.1) no depende de las variables de nivel superior; por el contrario, en (P) sí hay dependencia.

Se plantea la siguiente heurística para salvaguardar este hecho: sea \mathbf{x}^* una solución óptima encontrada en alguna iteración del algoritmo. Al resolver el subproblema, se obtiene un par $(\mathbf{y}^*, \mathbf{z}^*)$ considerando la matriz $\mathbf{P}(\mathbf{x}^*)$. En la siguiente iteración, se añaden las restricciones con \mathbf{z}^* y $\mathbf{P}(\mathbf{x}^*)$ como coeficientes. Se obtiene un nuevo \mathbf{x}^* y se repite el proceso.

El fin de esta heurística es imponer que el diseño a entregar sea compatible con las soluciones entregadas en las iteraciones anteriores. En concreto, si se han resuelto ℓ iteraciones del problema, entonces se tienen $\mathbf{z}^1, \dots, \mathbf{z}^\ell$ soluciones discretas del subproblema. La heurística fuerza a que un nuevo diseño sea compatible con estas soluciones \mathbf{z}^i .

Apéndice C

Verosimilitud y Matrices de Información del modelo EBA

En este apéndice se procede a calcular la log-verosimilitud y la matriz de información de Fisher del modelo EBA para dos y tres alternativas.

C.1. El caso de dos alternativas

Considérese un diseño $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ con T preguntas. Sea n_{hit} igual a 1 si el consumidor h escoge la alternativa i en la pregunta t y 0 en otro caso. Sea $\alpha > 0$ el vector de los pesos de importancia. La verosimilitud de α dado un consumidor h y \mathbf{x} es

$$L_h(\alpha | \mathbf{x}, \mathbf{n}_h) = \prod_{t=1}^T \prod_{i=1}^2 p(\mathbf{x}_{it} | \alpha, \mathbf{x})^{n_{hit}},$$

donde \mathbf{n}_h es una matriz que codifica las respuestas del consumidor h y p está dada por (2.3). Si se considera una muestra aleatoria de H consumidores la verosimilitud es

$$L(\alpha | \mathbf{x}, \mathbf{n}) = \prod_{h=1}^H L_h(\alpha | \mathbf{x}, \mathbf{n}_h) = \prod_{h=1}^H \prod_{t=1}^T \prod_{i=1}^2 p(\mathbf{x}_{it} | \alpha, \mathbf{x})^{n_{hit}} = \prod_{t=1}^T \prod_{i=1}^2 p(\mathbf{x}_{it} | \alpha, \mathbf{x})^{\sum_{h=1}^H n_{hit}},$$

donde la última igualdad se justifica pues la probabilidad no depende de h . La log-verosimilitud del modelo queda

$$\begin{aligned} \ell(\alpha | \mathbf{x}, \mathbf{n}) &= \ln(L(\alpha | \mathbf{x}, \mathbf{n})) \\ &= \sum_{t=1}^T \sum_{i=1}^2 \sum_{h=1}^H n_{hit} \ln(p(\mathbf{x}_{it} | \alpha, \mathbf{x})) \\ &= \sum_{t=1}^T \sum_{i=1}^2 \sum_{h=1}^H n_{hit} \left[\ln \left(\sum_{k \in A_{it}} \alpha_k \right) - \ln \left(\sum_{k \in A_{1t} \cup A_{2t}} \alpha_k \right) \right]. \end{aligned}$$

Maximizando la log-verosimilitud se obtiene un vector estimador $\hat{\alpha}$. La expresión no es cerrada, así que se utilizan métodos de optimización numérica para obtener dicho estimador.

Dado que a todos los consumidores se les muestra el mismo cuestionario, a la hora de comparar dos cuestionarios en términos de la matriz de información basta considerar un sólo consumidor, es decir, $H = 1$. La matriz de información de Fisher tiene la forma

$$(\mathcal{I}(\widehat{\boldsymbol{\alpha}}))_{jl} = -\mathbb{E} \left(\left. \frac{\partial^2 \ell}{\partial \alpha_j \partial \alpha_l}(\boldsymbol{\alpha} \mid \mathbf{x}, \mathbf{n}) \right|_{\boldsymbol{\alpha} = \widehat{\boldsymbol{\alpha}}} \right),$$

en donde la esperanza se toma con respecto a la variable aleatoria que representa los datos. Si $\mathbf{1}_X$ es la función indicadora del conjunto X entonces, derivando ℓ con respecto a α_l , se obtiene

$$\frac{\partial \ell}{\partial \alpha_l}(\boldsymbol{\alpha} \mid \mathbf{x}, \mathbf{n}) = \sum_{t=1}^T \sum_{i=1}^2 n_{hit} \left[\frac{\mathbf{1}_{A_{it}}(l)}{\sum_{k \in A_{it}} \alpha_k} - \frac{\mathbf{1}_{A_{1t} \cup A_{2t}}(l)}{\sum_{k \in A_{1t} \cup A_{2t}} \alpha_k} \right],$$

y derivando nuevamente con respecto a un atributo j se obtiene

$$\frac{\partial^2 \ell}{\partial \alpha_j \partial \alpha_l}(\boldsymbol{\alpha} \mid \mathbf{x}, \mathbf{n}) = \sum_{t=1}^T \sum_{i=1}^2 n_{hit} \left[\frac{\mathbf{1}_{A_{1t} \cup A_{2t}}(l) \mathbf{1}_{A_{1t} \cup A_{2t}}(j)}{(\sum_{k \in A_{1t} \cup A_{2t}} \alpha_k)^2} - \frac{\mathbf{1}_{A_{it}}(l) \mathbf{1}_{A_{it}}(j)}{(\sum_{k \in A_{it}} \alpha_k)^2} \right].$$

Nótese que si $j = l$ el producto de las indicatrices en la expresión anterior es redundante. Para finalizar sólo basta tomar esperanza en los términos anteriores, reemplazando el dato n_{hit} por la variable aleatoria correspondiente. Por linealidad de la esperanza sólo hay que reemplazar n_{hit} por $\mathbb{E}(n_{hit}) = p(\mathbf{x}_{it} \mid \widehat{\boldsymbol{\alpha}}, \mathbf{x})$.

La log-verosimilitud y las matrices están bien definidas pues toda suma del estilo $\sum_{k \in A_{it}} \alpha_k$ es estrictamente positiva. En efecto, A_{it} nunca es vacío, pues de lo contrario $\mathbf{x}_{1t} = \mathbf{x}_{2t}$, lo que no tiene sentido. Como A_{it} no es vacío, existe algún $j \in A_{it}$, por lo que $\sum_{k \in A_{it}} \alpha_k \geq \alpha_j > 0$.

C.2. El caso de tres alternativas

Como en la sección C.1, considérese un diseño $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ con T preguntas. Sea n_{hit} igual a 1 si el consumidor h escoge la alternativa i en la pregunta t y 0 en otro caso. Sea $\boldsymbol{\alpha} > \mathbf{0}$ el vector de los pesos de importancia. La verosimilitud de $\boldsymbol{\alpha}$ dado un consumidor h y \mathbf{x} es

$$L_h(\boldsymbol{\alpha} \mid \mathbf{x}, \mathbf{n}_h) = \prod_{t=1}^T \prod_{i=1}^3 p(\mathbf{x}_{it} \mid \boldsymbol{\alpha}, \mathbf{x})^{n_{hit}},$$

donde \mathbf{n}_h es una matriz que codifica las respuestas del consumidor h y p está dada por (3.2). Abusando de notación, en lo que sigue se entiende que si en una expresión aparecen los elementos $i, j, l \in \{1, 2, 3\}$, entonces todos son distintos entre sí. Además, se utilizan las probabilidades $p_{ij}(\mathbf{x}_{it})$ dadas por (3.3). Si se considera una muestra aleatoria de H consumidores la verosimilitud es

$$L(\boldsymbol{\alpha} \mid \mathbf{x}, \mathbf{n}) = \prod_{t=1}^T \prod_{i=1}^3 p(\mathbf{x}_{it} \mid \boldsymbol{\alpha}, \mathbf{x})^{\sum_{h=1}^H n_{hit}},$$

donde la última igualdad se justifica pues la probabilidad no depende de h . La log-verosimilitud del modelo queda

$$\ell(\boldsymbol{\alpha} \mid \mathbf{x}, \mathbf{n}) = \sum_{t=1}^T \sum_{i=1}^3 \sum_{h=1}^H n_{hit} \left[\ln \left(\sum_{k \in A_{it}} \alpha_k + \sum_{k \in A_{ijt}} \alpha_k p_{ij}(\mathbf{x}_{it}) + \sum_{k \in A_{ilt}} \alpha_k p_{il}(\mathbf{x}_{it}) \right) - \ln(Z_{3t}) \right].$$

Maximizando la log-verosimilitud se obtiene un vector estimador $\widehat{\boldsymbol{\alpha}}$. La expresión no es cerrada, así que se utilizan métodos de optimización numérica para obtener dicho estimador.

Se procede a calcular la matriz de información de Fisher derivando la log-verosimilitud dos veces. Si $\mathbf{1}_X$ es la función indicadora del conjunto X y omitiendo el argumento \mathbf{x}_{it} de las probabilidades p_{ij} por simplicidad, entonces, derivando ℓ con respecto a α_r , se obtiene

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha_r}(\boldsymbol{\alpha}) = \sum_{t,i} n_{hit} & \left[\frac{\mathbf{1}_{A_{it}}(r) + \mathbf{1}_{A_{ijt}}(r)p_{ij} + \frac{\partial p_{ij}}{\partial \alpha_r} \sum_{k \in A_{ijt}} \alpha_k + \mathbf{1}_{A_{ilt}}(r)p_{il} + \frac{\partial p_{il}}{\partial \alpha_r} \sum_{k \in A_{ilt}} \alpha_k}{\sum_{k \in A_{it}} \alpha_k + \sum_{k \in A_{ijt}} \alpha_k p_{ij} + \sum_{k \in A_{ilt}} \alpha_k p_{il}} - \dots \right. \\ & \left. \dots - \frac{\mathbf{1}_{A_{1t}}(r) + \mathbf{1}_{A_{2t}}(r) + \mathbf{1}_{A_{3t}}(r) + \mathbf{1}_{A_{12t}}(r) + \mathbf{1}_{A_{13t}}(r) + \mathbf{1}_{A_{23t}}(r)}{Z_{3t}} \right], \end{aligned}$$

y derivando nuevamente con respecto a un atributo s se obtiene

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \alpha_s \partial \alpha_r}(\boldsymbol{\alpha}) = \sum_{t,i} n_{hit} & \left[\frac{\mathbf{1}_{A_{ijt}}(r) \frac{\partial p_{ij}}{\partial \alpha_s} + \mathbf{1}_{A_{ijt}}(s) \frac{\partial p_{ij}}{\partial \alpha_r} + \frac{\partial^2 p_{ij}}{\partial \alpha_s \partial \alpha_r} \sum_{k \in A_{ijt}} \alpha_k}{\sum_{k \in A_{it}} \alpha_k + \sum_{k \in A_{ijt}} \alpha_k p_{ij} + \sum_{k \in A_{ilt}} \alpha_k p_{il}} + \dots \right. \\ & \dots + \frac{\mathbf{1}_{A_{ilt}}(r) \frac{\partial p_{il}}{\partial \alpha_s} + \mathbf{1}_{A_{ilt}}(s) \frac{\partial p_{il}}{\partial \alpha_r} + \frac{\partial^2 p_{il}}{\partial \alpha_s \partial \alpha_r} \sum_{k \in A_{ilt}} \alpha_k}{\sum_{k \in A_{it}} \alpha_k + \sum_{k \in A_{ijt}} \alpha_k p_{ij} + \sum_{k \in A_{ilt}} \alpha_k p_{il}} - \dots \\ & \dots - \left(\frac{\mathbf{1}_{A_{it}}(r) + \mathbf{1}_{A_{ijt}}(r)p_{ij} + \frac{\partial p_{ij}}{\partial \alpha_r} \sum_{k \in A_{ijt}} \alpha_k + \mathbf{1}_{A_{ilt}}(r)p_{il} + \frac{\partial p_{il}}{\partial \alpha_r} \sum_{k \in A_{ilt}} \alpha_k}{\sum_{k \in A_{it}} \alpha_k + \sum_{k \in A_{ijt}} \alpha_k p_{ij} + \sum_{k \in A_{ilt}} \alpha_k p_{il}} \right) \times \dots \\ & \dots \times \left(\frac{\mathbf{1}_{A_{it}}(s) + \mathbf{1}_{A_{ijt}}(s)p_{ij} + \frac{\partial p_{ij}}{\partial \alpha_s} \sum_{k \in A_{ijt}} \alpha_k + \mathbf{1}_{A_{ilt}}(s)p_{il} + \frac{\partial p_{il}}{\partial \alpha_s} \sum_{k \in A_{ilt}} \alpha_k}{\sum_{k \in A_{it}} \alpha_k + \sum_{k \in A_{ijt}} \alpha_k p_{ij} + \sum_{k \in A_{ilt}} \alpha_k p_{il}} \right) + \dots \\ & \dots + \left(\frac{\mathbf{1}_{A_{1t}}(r) + \mathbf{1}_{A_{2t}}(r) + \mathbf{1}_{A_{3t}}(r) + \mathbf{1}_{A_{12t}}(r) + \mathbf{1}_{A_{13t}}(r) + \mathbf{1}_{A_{23t}}(r)}{Z_{3t}} \right) \times \dots \\ & \left. \dots \times \left(\frac{\mathbf{1}_{A_{1t}}(s) + \mathbf{1}_{A_{2t}}(s) + \mathbf{1}_{A_{3t}}(s) + \mathbf{1}_{A_{12t}}(s) + \mathbf{1}_{A_{13t}}(s) + \mathbf{1}_{A_{23t}}(s)}{Z_{3t}} \right) \right]. \end{aligned}$$

Las derivadas merecen un poco de atención. Nótese que las probabilidades $p_{ij}(\mathbf{x}_{it})$ siempre están bien definidas. Por contradicción, supóngase que $\mathbf{x}_{it} \neq \mathbf{x}_{jt}$ y que el denominador de $p_{ij}(\mathbf{x}_{it})$ vale cero. Se sigue que

$$\begin{aligned} Z_{3t} - \sum_{k \in A_{ijt} \cup A_{ilt}} \alpha_k = 0 & \iff \sum_{k \in A_{it}} \alpha_k + \sum_{k \in A_{jt}} \alpha_k + \sum_{k \in A_{ilt}} \alpha_k + \sum_{k \in A_{jlt}} \alpha_k = 0 \\ & \iff A_{it} = A_{jt} = A_{ilt} = A_{jlt} = \emptyset, \end{aligned}$$

donde la segunda equivalencia se tiene pues los términos α_k son todos mayores a cero. Esto se traduce en que $\mathbf{x}_{it} = \mathbf{x}_{jt}$ y que \mathbf{x}_{lt} no posea atributos comunes ni con \mathbf{x}_{it} ni con \mathbf{x}_{jt} , lo que contradice la suposición inicial. Por ende, las probabilidades están bien definidas, y las derivadas valen

$$\frac{\partial p_{ij}}{\partial \alpha_s} = \begin{cases} \frac{(1 - p_{ij})(\mathbf{1}_{A_{it}}(s) + \mathbf{1}_{A_{ilt}}(s)) - p_{ij}(\mathbf{1}_{A_{jt}}(s) + \mathbf{1}_{A_{jlt}}(s))}{Z_{3t} - \sum_{k \in A_{ijt} \cup A_{lt}} \alpha_k}, & \text{si } \mathbf{x}_{it} \neq \mathbf{x}_{jt}, \\ 0, & \text{si } \mathbf{x}_{it} = \mathbf{x}_{jt}, \end{cases}$$

$$\frac{\partial^2 p_{ij}}{\partial \alpha_s \partial \alpha_r} = \begin{cases} \frac{(\mathbf{1}_{A_{jt}} + \mathbf{1}_{A_{jlt}})(s) \left[(2p_{ij} - 1)(\mathbf{1}_{A_{it}} + \mathbf{1}_{A_{ilt}}) + 2p_{ij}(\mathbf{1}_{A_{jt}} + \mathbf{1}_{A_{jlt}}) \right] (r)}{\left(Z_{3t} - \sum_{k \in A_{ijt} \cup A_{lt}} \alpha_k \right)^2} - \dots \\ \frac{(\mathbf{1}_{A_{it}} + \mathbf{1}_{A_{ilt}})(s) \left[(1 - 2p_{ij})(\mathbf{1}_{A_{jt}} + \mathbf{1}_{A_{jlt}}) + 2(1 - p_{ij})(\mathbf{1}_{A_{it}} + \mathbf{1}_{A_{ilt}}) \right] (r)}{\left(Z_{3t} - \sum_{k \in A_{ijt} \cup A_{lt}} \alpha_k \right)^2}, & \text{si } \mathbf{x}_{it} \neq \mathbf{x}_{jt}, \\ 0, & \text{si } \mathbf{x}_{it} = \mathbf{x}_{jt}. \end{cases}$$

Por último, por linealidad de la esperanza, se reemplaza n_{hit} por $\mathbb{E}(n_{hit}) = p(\mathbf{x}_{it} \mid \hat{\boldsymbol{\alpha}}, \mathbf{x})$.

Apéndice D

Ejemplos de diseño

D.1. Dos alternativas, caso 2×3^2

Cjto. de elección	Perfil	Diseño EBA, tasa de traslape = 57,14 %			Diseño compensatorio, tasa de traslape = 29,16 %		
		Atributos			Atributos		
		1	2	3	1	2	3
1	I	2	3	3	2	1	1
	II	2	2	3	2	1	3
2	I	2	2	3	2	2	1
	II	2	1	3	1	1	2
3	I	2	2	3	1	1	1
	II	2	1	2	1	1	1
4	I	2	1	3	1	3	1
	II	2	1	2	2	2	2
5	I	2	1	3	1	2	1
	II	1	1	2	2	1	1
6	I	1	1	3	2	3	1
	II	1	1	2	1	1	2
7	I	1	1	3	2	3	2
	II	1	1	1	1	2	3

D.2. Tres alternativas, caso $2 \times 3 \times 4$

Cjto. de elección	Perfil	Diseño EBA, tasa de traslape = 100 %			Diseño compensatorio, tasa de traslape = 50 %		
		Atributos			Atributos		
		1	2	3	1	2	3
1	I	2	3	4	2	1	1
	II	2	3	4	1	2	2
	III	2	3	3	2	3	3
2	I	2	3	4	2	1	3
	II	2	3	4	1	2	4
	III	1	3	2	2	3	1
3	I	1	3	4	2	3	2
	II	1	3	2	1	1	3
	III	1	3	2	2	2	4
4	I	1	3	2	1	3	2
	II	1	2	2	2	1	3
	III	1	2	2	1	2	4
5	I	1	3	2	2	1	4
	II	1	2	2	1	2	1
	III	1	1	1	2	3	2
6	I	1	3	2	1	1	1
	II	1	1	2	2	2	2
	III	1	1	1	1	3	3