



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

HACIA LA IMPLEMENTACIÓN DE UN MODELO DE ELECCIÓN DISCRETA  
BASADO EN EL PRINCIPIO DECISION FIELD THEORY

MEMORIA PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL INDUSTRIAL

TOMÁS BENJAMÍN NEGRETE VALENZUELA

PROFESOR GUÍA:  
CRISTIAN ANGELO GUEVARA CUE

MIEMBROS DE LA COMISIÓN:  
RICARDO MONTOYA MOREIRA  
ANDRÉS MUSALEM SAID

SANTIAGO DE CHILE

2019

RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL  
POR: TOMÁS BENJAMÍN NEGRETE VALENZUELA  
FECHA: 2019  
PROF. GUÍA: CRISTIAN ANGELO GUEVARA CUE

HACIA LA IMPLEMENTACIÓN DE UN MODELO DE ELECCIÓN DISCRETA  
BASADO EN EL PRINCIPIO DECISION FIELD THEORY

La modelación de elecciones discretas tiene diversas aplicaciones prácticas relevantes en varias áreas del conocimiento, incluyendo economía, marketing, medio ambiente, salud y transporte. El enfoque de modelación usual considera que las personas realizan sus elecciones maximizando la utilidad agregada que perciben de las alternativas. A diferencia de los modelos tradicionales, el enfoque Decision Field Theory (DFT) considera un enfoque más realista que asume que el proceso de deliberación, es decir, el tiempo y el orden en que se observan los atributos, tiene un efecto en la elección final que se realiza. Esta memoria contribuye en buscar una propuesta de modelo basado en el principio DFT, subsanando limitaciones de identificación e interpretación del modelo actual que utiliza este enfoque.

Un problema práctico para la implementación de un enfoque DFT como un modelo de elección discreta es que normalmente sólo se tiene información de los valores de los atributos de cada alternativa y lo que la población escogió, pero no el tiempo que la persona tomó en decidir, ni lo que observó para realizar esta decisión. Para dar cuenta de esta limitación, esta propuesta considera que estos datos se consideran como latentes, es decir, no observados. Para lograr esto, el modelo propuesto calcula todas las posibles observaciones que un individuo puede realizar construyendo una formulación con base matemática y probabilística, desarrollando así el modelo en su forma básica.

El modelo propuesto se estudia mediante una simulación de Monte Carlo con el fin de analizar el su comportamiento, su validez, y la capacidad de recuperar el valor de los parámetros. En segundo lugar, se realiza una comparación de la propuesta con otros modelos de la literatura en base a los datos generados. Finalmente se utilizan bases de datos reales para contrastar el modelo propuesto con métodos tradicionales de estimación, en especial con DFT.

Si bien el modelo corresponde a un prototipo que aún requiere ciertos ajustes y relajación de hipótesis en su formulación, éste entrega resultados consistentes con su formulación y con los supuestos que se toman, además de mostrar resultados prometedores en cuanto a ajuste a los datos y factibilidad de estimación y de interpretación. El aporte generado se resume en una forma más realista modelar cómo los individuos realizan sus elecciones, subsanando limitaciones de identificación e interpretación de modelos propuestos en la literatura.

*Dedicado a todas las personas que sólo por exigir una vida más digna y un país más justo fueron violentadas por el Estado de Chile en las manifestaciones de finales de 2019.*

# Agradecimientos

Es el término de una gran etapa de mi vida, en la cual aprendí y crecí considerablemente. Es por esto que no sólo es necesario agradecer a las personas envueltas en este último año de trabajo en esta memoria, sino que también a todas las personas que fueron parte de esta etapa que se despide.

Agradezco a mi profesor guía Angelo Guevara por el apoyo constante en la realización de esta memoria, por la paciencia, dedicación y confianza depositadas en mí y este trabajo.

A mis padres, Eliana y Mario, que se han esforzado por entregarme todo para tener una formación profesional integra. Por inculcarme valores que, espero, puedan ayudarme a ser un profesional que ayude a construir un mejor país.

A mis hermanas, Paola y Pamela, que se han preocupado en sobremanera de proteger y educar a su hermano menor. Siempre estaré agradecido.

A toda mi familia "Les tolchites", especialmente a mi abuela Flor Vega (Q.E.P.D), por siempre entregarme un lugar donde sentirme feliz.

A mis amigos del colegio, especialmente a Adrián, Andrés y Sebastián, que han sido un pilar fundamental de alegría en este proceso.

A todos mis amigos de la universidad, y especialmente a Ignacia y Alvaro, que han sido un apoyo emocional en muchos momentos de esta etapa.

A mi familia internacional dispersa por el mundo, especialmente a Inês, Rafael, Jonatan, Jordi, Nicolás, Nicolás y André, que a pesar de que la distancia nos separe, la amistad se mantiene intacta.

Por último, a mi gato Cú (Q.E.P.D), por tu constante e incondicional compañía en este proceso.

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Objetivos . . . . .	2
1.3. Metodología . . . . .	3
1.4. Estructura de la memoria . . . . .	3
<b>2. Marco Teórico</b>	<b>4</b>
2.1. Proceso de toma de decisiones . . . . .	4
2.2. El proceso de búsqueda de información . . . . .	6
2.3. Introducción a modelos de elección discreta . . . . .	7
2.4. Modelos probabilísticos estáticos . . . . .	8
2.4.1. Random Utility Maximization . . . . .	9
2.4.2. Otros modelos . . . . .	11
2.5. Modelos estocásticos dinámicos . . . . .	12
2.5.1. Decision Field Theory (DFT) . . . . .	13
<b>3. Formulación de un modelo DFT basado en el principio de Maximización de la Utilidad Aleatoria</b>	<b>24</b>
3.1. Ejemplo introductorio . . . . .	25
3.2. Modelo econométrico . . . . .	25
3.2.1. Toma de decisión en cada paso $t$ . . . . .	27
3.2.2. Función de utilidad . . . . .	28
3.2.3. Expresiones de las probabilidades . . . . .	30
3.2.4. Ejemplo . . . . .	34
3.3. Probabilidades latentes . . . . .	34
3.4. Estimación . . . . .	35
3.4.1. Expresión matemática de la estimación . . . . .	35
3.4.2. Métodos de estimación . . . . .	36
3.5. Codificación en R . . . . .	38

3.6. Diferencias teóricas con Decision Field Theory . . . . .	39
3.7. Resumen ecuaciones del modelo prototipo . . . . .	40
<b>4. Ilustración, análisis y validación del modelo propuesto mediante casos simulados y reales</b>	<b>43</b>
4.1. Evaluación y validez del modelo . . . . .	43
4.1.1. Función de utilidad . . . . .	43
4.1.2. Parámetro $\delta_t$ : tolerancia . . . . .	46
4.1.3. Parámetro $U_t^0$ : Opinión inicial de alternativas . . . . .	47
4.1.4. Probabilidad de observar atributos . . . . .	50
4.2. Análisis del método de estimación: Monte Carlo . . . . .	51
4.2.1. Datos . . . . .	51
4.2.2. Estimación con camino conocido . . . . .	52
4.2.3. Estimación con cantidad de pasos de elección conocida . . . . .	56
4.2.4. Estimación convencional: probabilidades latentes de caminos de búsqueda	59
4.2.5. Decisión de especificación del prototipo . . . . .	66
4.2.6. Recuperación de parámetros . . . . .	67
4.2.7. Comparación con otros modelos . . . . .	68
4.3. Estimación base de datos real: caso SwissMetro . . . . .	69
4.3.1. Comparación con DFT . . . . .	71
4.3.2. Comportamiento modelo propuesto . . . . .	73
<b>5. Conclusiones</b>	<b>75</b>
<b>A. Anexo: Código R ejemplo generación de datos</b>	<b>79</b>
<b>B. Anexo: Código R ejemplo estimación</b>	<b>88</b>
<b>Bibliografía</b>	<b>92</b>

# Índice de Tablas

2.1. Categorización de teorías de decisión (Busemeyer y Townsend, 1993) . . . . .	8
2.2. DFT: Ejemplo de matriz de atributos M . . . . .	16
2.3. DFT: Ejemplo de matriz de contraste C . . . . .	17
2.4. DFT: Ejemplo de matriz de distancia D . . . . .	17
2.5. DFT: Ejemplo de matriz de retroalimentación S . . . . .	18
3.1. Matrices para la estimación del modelo. Probabilidades y utilidades por caminos de observación . . . . .	39
3.2. Diferencias teóricas del modelo prototipo y Decision Field Theory . . . . .	39
4.1. Valor parámetros para generación de datos . . . . .	52
4.2. Puntos iniciales diferentes para estimación con parámetro $\alpha$ . . . . .	55
4.3. Estimación camino conocido para diferentes puntos iniciales con parámetro $\alpha$ . . . . .	55
4.4. Resultados de la estimación de los parámetros de una simulación para cantidad de pasos conocida y parámetro $\alpha$ desconocido . . . . .	59
4.5. Cantidad de veces que las alternativas fueron escogidas . . . . .	59
4.6. Puntos iniciales para estimación con factores de escala . . . . .	60
4.7. Resultados estimación con factores de escala para diferentes puntos iniciales . . . . .	61
4.8. Estimación parámetros para tiempos máximos de estimación diferentes. Siendo $T=7$ el tiempo máximo de elección de los individuos. . . . .	62
4.9. Desviaciones estándar estimación para Tiempos máximos diferentes . . . . .	63
4.10. Evaluación en cambios de los resultados de estimación al modificar el parámetro $\alpha$ , siendo $\alpha = 0,60$ el valor real con el que se generaron los datos . . . . .	64
4.11. Resultados estimación del modelo prototipo para datos generados con puntos iniciales heterogeneos . . . . .	65
4.12. Comparación métodos de estimación con parámetro $\alpha$ conocido, con cantidad de pasos máxima igual a 7 . . . . .	66
4.13. Comparación métodos de estimación con parámetro $\alpha$ estimado, con cantidad de pasos máxima igual a 7 . . . . .	67

4.14. Comparación métodos finales . . . . .	68
4.15. Resultados de la estimación de los parámetros con el modelo prototipo . . .	68
4.16. Comparación modelos con datos creados con el modelo propuesto . . . . .	69
4.17. Prueba de estimación del modelo prototipo con base de datos de Swissmetro mediante diferentes métodos de estimación . . . . .	70
4.18. Comparación modelos para la estimación de Swissmetro ( $N = 5607$ ), ultimas 4 columnas son calculadas por Belgiawan et al. (2017), escaladas en $\times 0,01$ .	71
4.19. Resultados modelo DFT para SwissMetro . . . . .	72
4.20. Comparación parámetros generales estimación SwissMetro entre Decision Field Theory y prototipo modelo . . . . .	72



# Índice de Ilustraciones

2.1. El proceso de toma de decisión (Engel et al., 1968) . . . . .	4
2.2. El proceso y contexto que lleva a una decisión (Ben-Akiva et al., 2012) . . .	6
2.3. El proceso de búsqueda de información (Xie, 2019) . . . . .	7
3.1. Decisiones involucradas en cada paso $t$ . . . . .	26
3.2. Probabilidades asociadas en cada paso $t$ . . . . .	28
3.3. Utilidades asociadas a la decisión de seguir buscando o no . . . . .	31
3.4. Generalización caminos posibles de un individuo . . . . .	35
4.1. Evolución utilidad de individuo para 3 parámetros con 3 atributos . . . . .	44
4.2. Efecto del valor del parámetro $\delta$ en el promedio de pasos totales de la población	47
4.3. Evolución de utilidades para mismo individuo con 3 casos diferentes (puntos iniciales trasladados) . . . . .	48
4.4. Evolución de utilidades para mismo individuo con 3 casos diferentes (puntos iniciales trasladados) con función utilidad futura modificada . . . . .	50
4.5. Valor media parámetros estimados de 20 simulaciones y valores reales (puntos azules) con $\alpha$ fijo . . . . .	53
4.6. t-test significancia a parámetros estimados de 20 simulaciones y nivel de sig- nificancia 0.05 (línea roja) . . . . .	54
4.7. Boxplot p-valor para test de media igual a valor real de 20 simulaciones. Línea roja es p-valor = 0,05 . . . . .	54
4.8. t-test significancia a parámetros estimados de 20 simulaciones y nivel de sig- nificancia 0.05 (línea roja) con $\alpha$ desconocido . . . . .	56
4.9. Boxplot p-valor para test de media igual a valor real de 20 simulaciones. Línea roja es p-valor = 0,05 . . . . .	57
4.10. Valor media parámetros estimados de 20 simulaciones y valores reales (puntos azules) . . . . .	58
4.11. Cantidad de pasos realizados de la población . . . . .	60

4.12. Simulación comportamiento de la utilidad de un individuo con parámetros estimados (Swissmetro), con atributos observados en cada paso: T= tiempo de viaje y C= costo de viaje . . . . .	73
4.13. Simulación comportamiento de la utilidad de una alternativa de 40 individuos con parámetros estimados (Swissmetro) . . . . .	74

# Capítulo 1

## Introducción

### 1.1. Motivación

En el área de la psicología, se define el proceso de toma de decisiones como el proceso cognitivo por el cual los individuos realizan la selección de alguna alternativa entre un número finito de éstas, buscando solucionar un problema propio. Estas decisiones se realizan en base a los valores de estas alternativas, en las preferencias y las creencias del tomador de decisión.

Actualmente, las formas más utilizadas para poder explicar cómo los individuos realizan sus elecciones consideran un enfoque completamente compensatorio (McFadden, 1976; Thurstone, 1927). En este tipo de enfoques se ve al tomador de decisión como un individuo que considera todos los atributos de todas las alternativas en el mismo instante y balancea la utilidad entregada por los atributos al tomar una decisión. Ejemplos de este enfoque son los Modelos de Utilidad aleatoria, como Multinomial Logit, Probit, Logit, entre otros modelos, en que las utilidades se ven de la forma (generalmente):

$$U_{nit} = \alpha_i + \sum_i^I \beta_{it} \cdot X_{nit} + \varepsilon_{nit} \quad (1.1)$$

El enfoque compensatorio de modelación de elección parece poco realista, desde un punto de vista psicológico, debido a que en su formulación no se considera el proceso cognitivo que los individuos realizan al tomar una decisión y asume que estos pueden analizar el mercado por completo de manera instantánea y objetiva.

En cambio, el enfoque DFT (Decision Field Theory) propone un proceso de elección que

parece más realista. Este modelo asume que el individuo va creando su utilidad a través del tiempo, en donde en cada instante  $t$ , el individuo observa sólo una característica a la vez, compara entre alternativas y luego, en base a esto 'actualiza' la percepción global que tiene de cada alternativa (Busemeyer y Townsend, 1993).

Existe evidencia que sugiere que el modelo de Decision Field Theory (DFT) se ajusta satisfactoriamente a los datos, y en algunos casos, incluso mejor que los modelos convencionales (Hancock et al., 2018). Sin embargo, este modelo tiene la limitación de que utiliza un enfoque basado en matrices que dificulta su estimación, identificación de parámetros, pronóstico e interpretación. Estas limitaciones están descritas con mayor profundidad en la sección de Marco Teórico, en la Sección 2.5.1.6.

Esta memoria busca entregar una nueva forma de poder explicar el comportamiento de los individuos al elegir una única opción entre alternativas, formulando un prototipo de un modelo que incluya la base de modelos convencionales de maximización de utilidad aleatoria y la lógica detrás de Decision Field Theory.

## 1.2. Objetivos

El objetivo general de esta memoria es formular un prototipo de un modelo de elección discreta que pueda explicar las elecciones de los individuos en base a una construcción dinámica de valores de preferencia, a partir de una evaluación secuencial de cada atributo. Este nuevo enfoque buscará evitar las limitaciones actuales del modelo DFT.

Luego, los objetivos específicos de esta memoria son:

1. Formular las bases de un modelo dinámico basado en los modelos MNL y DFT, que pueda representar la evolución de utilidad de los individuos, considerando una implementación y configuración del modelo más simple que DFT.
2. Definir un método de estimación del modelo formulado.
3. Establecer y aplicar medidas de desempeño para el modelo formulado, comparándolo con los modelos actuales utilizados en la literatura, especialmente con DFT.
4. Analizar los resultados para identificar y caracterizar los principales beneficios y principales falencias del modelo, y así establecer bases para su continuación y mejora.

## 1.3. Metodología

A grandes rasgos, la metodología utilizada para realizar la investigación de esta memoria se divide en 4 etapas. Primero, se realizó un estudio para la formulación del prototipo propuesto. Luego, se estudió el método de estimación, al realizar esto, se modificó constantemente el modelo para poder tener una formulación y método de estimación coherentes. En tercer lugar, se estudió la validez y el comportamiento del modelo propuesto, generando una simulación de Monte Carlo para ello. Finalmente, se realizó una comparación con otros modelos de la literatura, utilizando la base de datos simulada, y también bases de datos reales.

Todas las etapas recientemente mencionadas fueron, en cierta medida, cíclicas, es decir, se formulaba la base de un modelo, se estudiaba su estimación y se validaba, si se encontraban posibles mejoras, se cambiaba la base matemática y se procedía nuevamente a estudiarla. Todo lo anterior hasta lograr llegar al prototipo que se presenta en este documento.

## 1.4. Estructura de la memoria

El presente documento se divide en 4 capítulos principales, los cuales serán introducidos en esta sección:

- Marco Teórico: En este capítulo, se introduce al lector el área del conocimiento en que esta memoria se envuelve. Se explican, a modo resumen, las metodologías ya existentes. Se muestran, además, los supuestos en los que se basa esta investigación.
- Formulación: El Capítulo 3 introduce y explica el modelo prototipo propuesto en esta memoria, entregando la base matemática completa del modelo, además de los supuestos que éste realiza.
- Evaluación: En el capítulo de evaluación se realiza un diagnóstico del modelo, en donde se observa el comportamiento de éste y la validez de los supuestos realizados. Además, se estudian algunos métodos de estimación, para comprobar si se pueden recuperar los parámetros, y si existen falencias en su formulación. Finalmente, se compara el prototipo con otros modelos de la literatura.
- Conclusiones: Se concluye respecto a las fortalezas y posibles mejoras del modelo. Además, se entregan ciertas directrices de cómo mejorar las falencias que éste tiene.

# Capítulo 2

## Marco Teórico

### 2.1. Proceso de toma de decisiones

En el área de la psicología, se define el proceso de toma de decisiones como el proceso cognitivo por el cual los individuos realizan la selección de alguna alternativa entre un número finito de éstas, buscando solucionar un problema propio. Estas decisiones se realizan en base a los valores de estas alternativas, en las preferencias y las creencias del tomador de decisión.

Para poder describir este proceso, (Engel et al., 1968) lo divide en 5 subprocesos, los cuales forman las etapas generales que un individuo (teóricamente) realiza para realizar la acción de elegir. Estas etapas se muestran en la Figura 2.1.

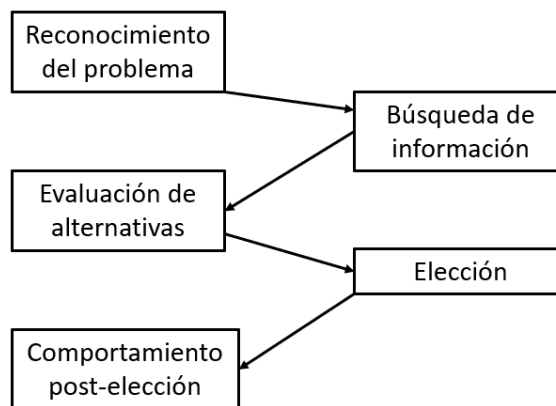


Figura 2.1: El proceso de toma de decisión (Engel et al., 1968)

De forma general, el tomador de decisión debe reconocer un problema, buscar información

(de su memoria o en fuentes externas), analizar sus opciones y realizar la elección para solucionar su problemática. De forma más específica, se describe el proceso a continuación:

1. Reconocimiento del problema: Darse cuenta de que algo no está como debería, o que algo podría mejorarse, por lo tanto, se detecta una problemática que debe ser solucionada.
2. Búsqueda de información: Se busca información que pueda solucionar el problema específico. Existen dos formas en que el individuo realiza su búsqueda de información.
  - (a) Búsqueda interna: El individuo busca información en su memoria y sus experiencias pasadas. Por lo tanto, esta búsqueda requiere menor esfuerzo para el individuo.
  - (b) Búsqueda externa: Se busca información de posible solución en fuentes externas. Luego, se asume que el individuo está realizando una decisión que requiere más esfuerzo.
3. Evaluación de alternativas: Se evalúan todas las alternativas factibles que el individuo necesita para la resolución de su problemática, esta evaluación en base a características visibles y que el tomador de decisión tiene acceso.
4. Elección: Se decide la alternativa que será utilizada para la solución del problema.
5. Comportamiento post elección

Si bien esta representación del modelo es altamente aceptada, existen postulaciones que buscan explicar de forma más específica los sub-procesos que ocurren en este proceso de toma de decisión. Por ejemplo, McFadden (1999) busca incluir las creencias, percepciones y las actitudes de los individuos que afectan en su decisión final. Ben-Akiva et al. (2012) especifican aún más el modelo anterior, incluyendo el contexto social y de mercado que tiene el individuo, como se puede apreciar en la Figura 2.2. En esta representación, se asume que el individuo tiene un contexto propio que afecta en cómo realiza sus decisiones.

Todas las representaciones del proceso de toma de decisión concuerdan en que los tomadores de decisión deben pasar por la etapa de búsqueda de información para realizar su propósito. Este paso es donde se basa el área del estudio de esta memoria, especialmente tratar de incluir la búsqueda externa, en donde el individuo adquiere su información y la procesa.

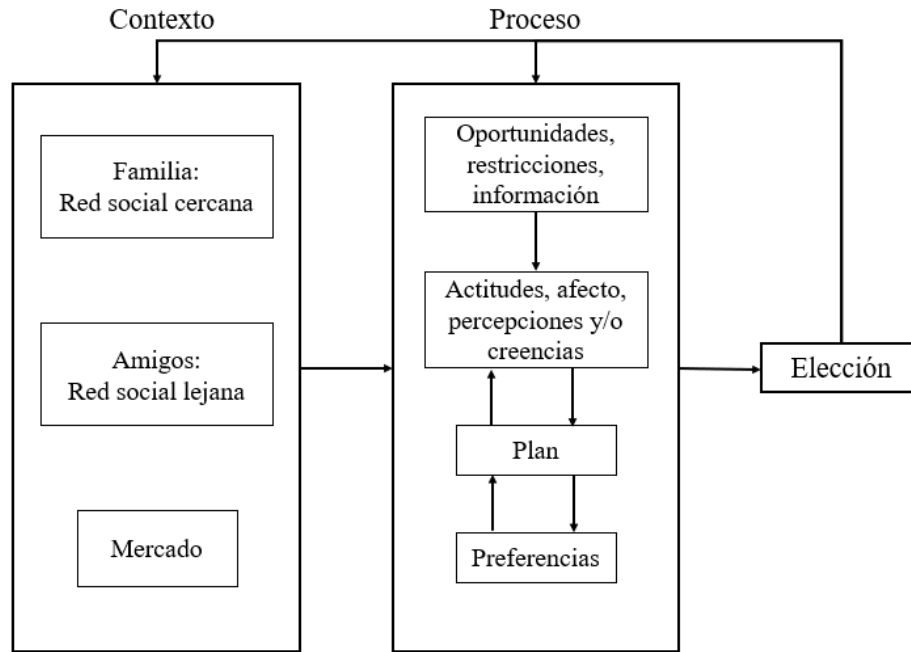


Figura 2.2: El proceso y contexto que lleva a una decisión (Ben-Akiva et al., 2012)

## 2.2. El proceso de búsqueda de información

La búsqueda de información es el proceso que describe cómo el individuo indaga sobre lo necesario para realizar su elección. Este proceso considera la búsqueda interna (memoria y experiencias previas) y la búsqueda externa, que es la adquisición y procesamiento de la información (Hulland y Kleinmuntz, 1994; Lynch y K. Srull, 1982).

El esfuerzo que se utiliza en la búsqueda interna y externa depende del contexto de elección, como el costo de conseguir la información, la presión en el tiempo, los incentivos para elegir correctamente, y la cantidad de experiencia previa que tiene el individuo (Hulland y Kleinmuntz, 1994; Payne et al., 1993). Durante esta memoria se realizarán solo supuestos respecto a la experiencia previa de los individuos (búsqueda interna) y no se entrará en mayor detalle.

Esta memoria se enfoca en el proceso de búsqueda externa, en tratar de incluir en la modelación matemática el cómo el individuo procesa la información que va adquiriendo. Siendo el procesamiento de información de vital importancia, debido a que el trabajo que se presentará durante todo este documento, se enfoca en incluir el proceso cognitivo previo a la decisión, sin tener información de éste. En la Figura 2.3 se puede observar los sub-procesos involucrados en el proceso de búsqueda de información, en donde se muestra el sub-proceso que se quiere estudiar (búsqueda externa) y cómo, teóricamente, se descompone lo que ocurre



en la mente de los individuos.

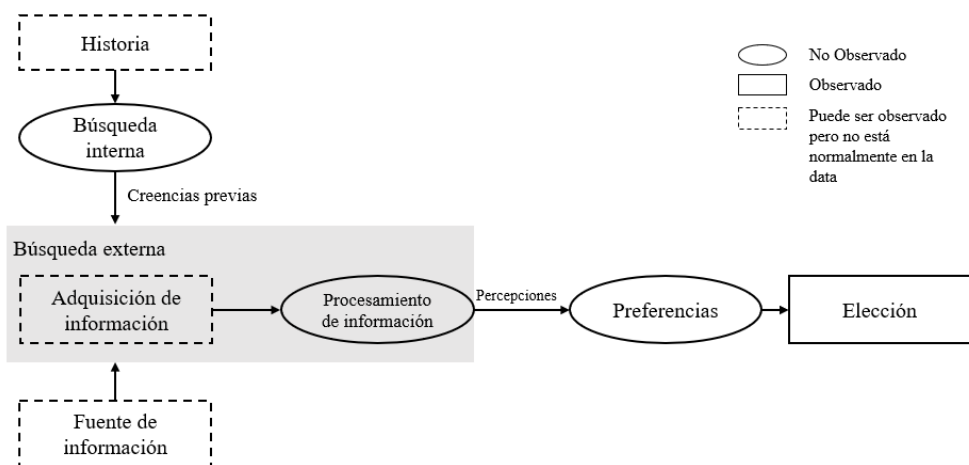


Figura 2.3: El proceso de búsqueda de información (Xie, 2019)

Para poder explicar matemáticamente cómo los individuos realizan sus elecciones nacen los modelos de elecciones discretas, que utilizan la información disponible (elección realizada y características visibles de las alternativas) para esto.

## 2.3. Introducción a modelos de elección discreta

Los modelos de elección discreta son modelos matemáticos que buscan describir, explicar y predecir la elección de un conjunto de alternativas basados en características (atributos) de éstas. Todo esto, con el objetivo de poder estudiar el comportamiento de elección de los individuos, y por lo tanto, realizar estudios de demanda y planificación. Los supuestos básicos de esta disciplina son:

- Existe un tomador de decisión (individuo, grupo, hogar, etc)
- Este tomador de decisión tiene características propias (a veces consideradas en la modelación)
- El individuo escoge una alternativa de un conjunto finito de éstas
- Estas alternativas tienen atributos que las describen
- Para la modelación, existe una regla de decisión (utilidad subyacente, preferencia, satisfacción, etc.)

En resumen, los modelos de elecciones discretas buscan poder explicar cómo los individuos

realizan sus elecciones teniendo disponible solamente información limitada de las alternativas y de los individuos.

Es importante destacar, que para esta memoria, el lenguaje que se utilizará es el ocupado en esta área del conocimiento. Es decir, se utilizará en gran parte las palabras *alternativa*, *atributo*, *elección*, *individuo*, *modelo* y los verbos *escoger* y *elegir*.

Existen diversos tipos de modelación de elecciones, los cuales pueden ser divididos en 4 categorías, como se muestra en la Tabla 2.1.

Los modelos determinísticos postulan que existe una relación binaria entre las alternativas, es decir, que dados ciertos valores de los atributos de las alternativas, una alternativa siempre será elegida si se repiten esas condiciones. Por otro lado, los modelos probabilísticos asumen que existe una probabilidad asociada a la elección de una alternativa, por lo tanto, para mismos valores de atributos, puede que se escoja otra alternativa en diferentes circunstancias.

Categoría	Estático	Dinámico
Determinísticos	Utilidad esperada	Dinámicas de acción
Probabilísticos	Utilidad aleatoria	Decision Field Theory

Tabla 2.1: Categorización de teorías de decisión (Busemeyer y Townsend, 1993)

Los modelos estáticos suponen que la preferencia o la probabilidad es independiente del tiempo de deliberación. El tiempo de deliberación es el tiempo que le toma al individuo realizar su elección desde que comienza su proceso de adquisición de ésta. Luego, los modelos dinámicos incorporan en su formulación que la probabilidad o la preferencia va cambiando a través del tiempo (Busemeyer y Townsend, 1993).

Existen diversos tipos de modelos para cada categoría, pero en la tabla solamente se encuentran los modelos más representativos de éstas. Para términos de esta memoria, se ignorará los modelos determinísticos, ya que se ha comprobado que son pocos realistas (Camerer, 1989; Edwards, 1955; Mosteller y Noguee, 1951).

## 2.4. Modelos probabilísticos estáticos

Los modelos probabilísticos estáticos son los modelos que, según Busemeyer y Townsend (1993), no consideran el proceso de deliberación (proceso cognitivo) que realizan los individuos al momento de encontrarse en una situación de elección. Además, al ser estocásticos,

consideran que existe una probabilidad asociada para cada alternativa, definida por un criterio que depende del modelo específico.

Normalmente, estos modelos requieren información de los atributos de las alternativas, la elección que realizó el individuo en cada situación de decisión, y, dependiendo del modelo y del analista, información del individuo.

Las ventajas de los modelos estocásticos estáticos, es su simplicidad y poder explicativo dentro de la gama de modelos existentes. El modelo Logit (McFadden, 1976) es el modelo más utilizado en la literatura, siendo un modelo fácil de aplicar y con buen poder explicativo. Existen diversas mejoras al modelo, que ayudan a mejorar aún más su calidad.

La gran desventaja de estos modelos, es que no consideran el proceso de deliberación mencionado anteriormente, al menos, no explícitamente en la formulación de éstos.

### **2.4.1. Random Utility Maximization**

El método de maximización de utilidad aleatoria (RUM) supone que el tomador de decisión tiene una capacidad perfecta de discriminación. Sin embargo, se asume que el analista no tiene toda la información utilizada por el tomador de decisión, por lo tanto existe una aleatoriedad en los cálculos matemáticos de la información que se tiene disponible.

Luego, se asume que sólo con conocer la información de los atributos de las alternativas y la elección realizada, se puede obtener una probabilidad de elección para cada una de estas alternativas, dependiendo de una utilidad aleatoria no observable que obtienen los individuos. Es decir, cada individuo tiene una apreciación diferente de cada alternativa, y la combinación matemática de los atributos la crea. Sin embargo, como se mencionó, esta utilidad calculada por el analista es imperfecta, por lo tanto, para determinar si la utilidad entregada por alternativa es mayor a otra, se calcula la probabilidad de que esto sea verdad.

A diferencia de estos modelos, los modelos determinísticos asumen que siempre se escogerá la opción con mayor utilidad sin considerar que existe una probabilidad de que eso no sea cierto. Los modelos RUM fueron primeramente estudiados y postulados por Thurstone (1927).

En términos generales, se define la utilidad subyacente aleatoria en función de los valores de los atributos de las alternativas estudiadas. Se multiplica el valor de los atributos por un parámetro que captura el efecto y la magnitud que éste tiene. En la Ecuación 2.1 se aprecia la expresión general de la utilidad de un individuo, donde  $X_{ni}$  es el vector de atributos de la alternativa  $i$ ,  $\beta_i$  es el vector del efecto de cada atributo,  $\alpha_i$  es la constante específica de alternativa, que captura diferencias intrínsecas entre las alternativas, y ,finalmente,  $\varepsilon_{ni}$  es el

error que puede se puede dar debido al conocimiento imperfecto del analista.

$$U_{ni} = \alpha_i + \sum_i^I \beta_i \cdot X_{ni} + \varepsilon_{ni} \quad (2.1)$$

Luego, la utilidad se divide en una parte sistemática (determinística) y una parte aleatoria (Ecuación 2.2).

$$\underbrace{U_{ni}}_{\text{Utilidad aleatoria}} = \underbrace{V_{ni}}_{\text{Utilidad sistemática}} + \underbrace{\varepsilon_{ni}}_{\text{Error}} \quad (2.2)$$

Con esto, se puede calcular la probabilidad de que la utilidad de una alternativa sea mayor a otra para un caso binario (Ecuación 2.3) y para el caso general (Ecuación 2.4)

$$\begin{aligned} \mathbb{P}(1) &= \mathbb{P}(U_{n1} \geq U_{n2}) \\ &= \mathbb{P}(U_{n1} - U_{n2} \geq 0) \\ &= \mathbb{P}(\varepsilon_{n2} - \varepsilon_{n1} \leq V_{n1} - V_{n2}) \\ &= F_{\varepsilon_{n2} - \varepsilon_{n1}}(V_{n1} - V_{n2}) \end{aligned} \quad (2.3)$$

$F_{\varepsilon_{n2} - \varepsilon_{n1}}$  es la función de distribución acumulada de  $\varepsilon_{n2} - \varepsilon_{n1}$  (univariada).

$$\mathbb{P}(i|C_n) = \mathbb{P}(\varepsilon_{in} - \varepsilon_{jn} \leq V_{in} - V_{jn} \forall j \in C_n) \quad (2.4)$$

Además, diferentes supuestos sobre la distribución de los errores  $\varepsilon$  resultan en diferentes modelos. Por ejemplo, si se asume que el error  $\varepsilon$  distribuye normal  $N(0, \Sigma)$  se obtiene el modelo Probit (Ecuación 2.5). El problema del modelo Probit es que su expresión de utilidad es muy compleja, es decir, implica el uso de integrales y, por cada alternativa se agrega una integral para su cálculo, lo que la vuelve computacionalmente exigente y de mayor complejidad para su configuración.

$$\mathbb{P}(i|C_n) = \int_{-\infty}^{V_{1n} - V_{2n}} \int_{-\infty}^{V_{1n} - V_{3n}} \dots \int_{-\infty}^{V_{1n} - V_{C_n n}} n(q; 0, \sigma_1) dq \quad (2.5)$$

En cambio, al asumir que el error distribuye Valor extremo tipo 1  $(0, \Sigma)$  la expresión final de probabilidad no depende de integrales, sino que es una expresión cerrada y más simple, como se puede ver en la Ecuación 2.6. Este modelo se conoce como el modelo Multinomial Logit (MNL) (McFadden, 1976).

$$\mathbb{P}(i|C_n) = \frac{e^{\mu V_{in}}}{\sum_j e^{\mu V_{jn}}} \quad (2.6)$$

Luego, para poder calcular y estimar los parámetros (efectos de alternativas) de este modelo utilizando estas probabilidades, se debe calcular la verosimilitud de que se reproduzcan los eventos (elecciones). Definiendo la elección  $y_{ni}$  hecha por el individuo  $n$  de la alternativa  $i$  como:

$$y_{ni} = \begin{cases} 1 & \text{si individuo } n \text{ eligió la alternativa } i \\ 0 & \text{si no la escogió} \end{cases}$$

Se tiene que la verosimilitud de observar la información se escribe como:

$$L = \prod_n^N \prod_i^I \mathbb{P}(i)^{y_{ni}} \quad (2.7)$$

Finalmente, esta verosimilitud Ecuación 2.7 debe ser optimizada para obtener los parámetros del modelo que mejor reproduzcan las probabilidades de los eventos.

Dentro de las ventajas del modelo MNL se encuentra que la distribución valor extremo se aproxima apropiadamente a la distribución normal, por lo que el poder explicativo que se pierde no es mayor (suponiendo que el error distribuye normal realmente). Además de ser el modelo más utilizado, es el modelo más estudiado y mejorado, con muchas variantes, especialmente el modelo Mixed Logit (McFadden y Train, 2000), el cual considera que todos los parámetros son diferentes para cada individuo, lo que mejora considerablemente el poder explicativo, ya que es más realista que el MNL simple. Otra mejora considerable es el Nested Logit (Ben-Akiva, 1973; Ben-Akiva y Lerman, 1974), que asume que puede existir una alta correlación entre alternativas, y por lo tanto, captura ese efecto.

Por último, una gran ventaja de este modelo es que puede capturar el valor de la disposición a pagar por cada característica, entre ellas, el valor del tiempo. Además, el modelo es el más utilizado, y hasta ahora, debido a su masificación, el más simple de aplicar.

### 2.4.2. Otros modelos

El modelo de minimización de arrepentimiento aleatorio (Random Regret Minimization), propuesto por Chorus (2010) y Chorus et al. (2008), asume que cuando los individuos realizan elecciones, su prioridad es minimizar su arrepentimiento luego del evento, en vez de maximizar su utilidad. Este modelo es tan simple, en términos de configuración, como el

RUM y, además, puede explicar el compromise effect, lo que los modelos RUM han ignorado (Chorus y Bierlaire, 2013).

Relative Advantage Model (RAM) o el modelo de ventaja relativa, es un modelo propuesto por Tversky y Simonson (1993) y, renombrado y mejorado por Kivetz et al. (2004). Este modelo nace debido a los intentos de explicar de mejor manera el “compromise effect” (Leong y Hensher, 2015) que el modelo tradicional RUM. El modelo es bastante similar a RUM, sin embargo, una diferencia fundamental es que el modelo RAM busca incluir explícitamente las desventajas y ventajas de cada alternativa, incluyendo un ratio de ventajas  $v/s$  desventajas en su formulación matemática Leong y Hensher (2015).

No se explicarán en detalle los modelos RAM y RRM, ya que sólo se utiliza en este documento como comparación del modelo prototipo para la base de datos Swissmetro en la sección 4.3. Para esta base de datos, estos modelos fueron calculados por Belgiawan et al. (2017).

## 2.5. Modelos estocásticos dinámicos

En la categorización definida por Busemeyer y Townsend (1993) los modelos dinámicos consideran que el tiempo que le toma al individuo decidir sí afecta en la decisión final que este hará. Dashiell (1937) descubrió que el tiempo de elección es una función decreciente de la magnitud de la fuerza de una preferencia. Jamieson y Petrusic (1977) concluyeron que el tiempo de decisión puede ser una forma eficiente de estimar la fuerza de la preferencia. Otros autores también relacionaron e investigaron el efecto del proceso de deliberación con la elección final (Ben Zur y Breznitz, 1981; Busemeyer, 1985; Wallsten y Barton, 1982).

Busemeyer y Townsend (1993) argumentan que debido a que aparentemente todas las decisiones involucran presión en el tiempo, los tomadores de decisión frecuentemente limitan su tiempo de elección. Esto podría ser la razón de que existen decisiones que a través de otros modelos se considerarían irracionales, por lo tanto, si se entregan incentivos apropiados para el proceso de decisión, los individuos finalmente escogerían las opciones que se consideran racionales. Luego, a pesar de que no exista tiempo límite establecido, los individuos se auto-definen implícitamente uno. Payne et al. (1992) realizaron un experimento y concluyeron que los tomadores de decisión sí realizan un trade-off entre precisión (elegir una alternativa que solucione el problema de mejor forma) y esfuerzo (tiempo de decisión).

Entonces, se debe tener un modelo que atrape el hecho de que las decisiones toman tiempo, y que la cantidad de tiempo que se utiliza influencia en la decisión final. Los modelos estáticos

no consideran explícitamente esta lógica, incluyendo los modelos RUM, que son los más utilizados para estudiar a los individuos.

### 2.5.1. Decision Field Theory (DFT)

Decision Field Theory (DFT) es un acercamiento dinámico a cómo los individuos toman sus decisiones. A diferencia de los modelos tradicionales, este modelo busca describir el proceso cognitivo por el cual los individuos realizan sus elecciones siguiendo la dinamicidad del mismo.

El modelo asume que los individuos realizan una búsqueda en anchura (breadth-first), es decir, que en cada momento de observación, se enfoca la atención en solamente un atributo y se comparan todas las alternativas, es decir, se observa atributo por atributo. A pesar de que existen otras formas de búsqueda, por ejemplo, alternativa por alternativa, y un híbrido entre los dos casos anteriores (Montgomery y Svenson, 1976; Payne, 1976), Busemeyer y Townsend (1993) deciden utilizar el protocolo de atributo por atributo, ya que se considera más acorde para los casos en su investigación. Luego, años más tarde, Noguchi y Stewart (2014) analizaron mediante tecnología Eye-tracking que las alternativas son comparadas (atributo por atributo) en vez de ser individualmente analizadas (alternativa por alternativa), para elecciones con 2 alternativas, por lo tanto, la hipótesis de Busemeyer y Townsend (1993) es correcta, al menos para estos casos.

Existen 2 criterios para realizar la elección en este modelo, el primero es cuando el individuo llega a un umbral interno para una de las alternativas, y por lo tanto, elige. El segundo, es cuando el individuo alcanza un umbral externo, como por ejemplo, el tiempo de respuesta límite. Esto se debe a que existirían dos formas de decisión posibles. La primera señala que algunos individuos eligen alternativas cuando son suficientemente buenas (satisficing behaviour, Simon, 1957), mientras que la segunda sugiere que otros utilizan todo el tiempo disponible y eligen de acuerdo a lo que pueden observar en este tiempo (maximising behaviour, Schwartz et al., 2002).

Diversos estudios, mejoras y comparaciones se han hecho del modelo, en donde se ha avanzado en su entendimiento y capacidad, siendo Hancock et al. (2018) los últimos en modificar el modelo (otras modificaciones considerables fueron propuestas por Berkowitsch et al., 2014; Hey et al., 2010; Roe et al., 2001).

### 2.5.1.1. Ventajas

Las ventajas del modelo de Decision Field Theory son diversas, además considerar explícitamente el proceso de deliberación en su formulación. Algunas de éstas, se han reportado, bajo ciertos supuestos:

1. Diederich, 1997 argumenta que DFT puede explicar fenómenos como "preference reversal"(reversión de preferencias Lichtenstein y Slovic, 1973), en cambio los modelos estáticos no pueden realizar esto.
2. Estudiando individuos mediante eye-tracking, se observa que generalmente se realiza comparación breadth-first (atributo por atributo) cuando solamente se comparan 2 alternativas, por lo que el modelo es apropiado para, al menos, estos casos de 2 opciones (Noguchi y Stewart, 2014). Con más de 2 alternativas aún no se realiza un estudio.
3. DFT predijo precisamente el porcentaje de participantes y sus elecciones, dado conocido el tiempo que le tomó realizar la decisión a cada individuo (Qin et al., 2013)
4. Se ha comprobado, que bajo ciertos contextos, DFT ajusta y predice mejor que otros modelos tradicionales, como MNL y RRM (Hancock et al., 2018).

### 2.5.1.2. Mecanismos básicos

Los mecanismos básicos del modelo DFT formulado por Busemeyer y Townsend (1993), y modificado (última modificación) por Hancock et al. (2018), se explicarán a continuación, a modo de entender su funcionamiento y el porqué de sus limitaciones. Todo lo mencionado a continuación se basa en los artículos recién mencionados.

Es importante destacar, que en este modelo, los atributos pueden ser visitados más de una vez, y cada atributo puede ser visitado consecutivamente. Esto se debe a que Busemeyer y Townsend (1993) considera que los pasos de observación representan la cantidad de tiempo que toma el obtener y procesar la información. En otras palabras, el proceso de deliberación  $T$  es igual a la cantidad de pasos  $t$  por la duración de los pasos  $t$ .

Cada alternativa tiene un valor de preferencia que cambia en el tiempo, la cual puede ser considerada como análoga a la utilidad subyacente del modelo RUM. Se define la columna de la matriz  $P_t$  como el vector que contiene el valor de preferencia de todas las alternativas en cada momento  $t$  de observación para un individuo. Los instantes  $t$  son definidos como cada paso en que el individuo observa un atributo de todas las alternativas y las compara. En la Ecuación 2.8 se observa la expresión principal de cómo se obtiene el valor de las preferencias



para cada instante  $t$ , como función de las preferencias en el instante  $t - 1$  y del atributo que se revisa en el instante  $t$ .

$$P_t = S \cdot P_{t-1} + V_t \quad (2.8)$$

En donde  $V_t$  (Ecuación 2.9) es el vector de valencia, el cual contiene el valor que entregan las alternativas dado el atributo que se observa en el instante  $t$ .  $C$  es la matriz de contraste donde  $c_{i,i} = 1$  y  $c_{j \neq i} = (-1)/(n - 1)$ , donde  $n$  es el número total de alternativas.  $M$  es la matriz de atributos, la cual contiene los valores de los atributos.  $W_t$  es el vector que indica qué atributo se observa en el instante  $t$ , lo que es definido por la probabilidad de observar el atributo  $w_i$ . Luego  $\varepsilon_t$  es el vector de errores, los cuales usualmente distribuyen  $\varepsilon \sim N(0, 1)$ .

$$V_t = C \cdot M \cdot W_t + \varepsilon_t \quad (2.9)$$

$C$  es la matriz de contraste donde  $c_{i,i} = 1$  y  $c_{j \neq i} = (-1)/(n - 1)$ , donde  $n$  es el número total de alternativas.  $M$  la matriz de atributos, la cual contiene los valores de los atributos.  $W_t$  es el vector que indica qué atributo se observa en el instante  $t$ . Luego  $\varepsilon_t$  es el vector de errores, los cuales usualmente distribuyen  $\varepsilon \sim N(0, 1)$ .

$S$ , en la Ecuación 2.8 es un vector de “peso del pasado”, el cual multiplica el valor de las preferencias del instante anterior, es decir, la preferencia que se había guardado hasta ese entonces. El cálculo de este vector se encuentra en la Ecuación 2.10

$$S = I - \phi_2 * e^{-\phi_1 \cdot D^2} \quad (2.10)$$

En donde  $\phi_1$  es un parámetro de sensibilidad (“¿qué tanto compiten alternativas similares?”) y  $\phi_2$  un parámetro de memoria (“¿es la información inicial o tardía más importante?”).  $I$  es la matriz de identidad y  $D^2$  es un vector que compara los valores de los atributos de acuerdo al criterio definido de antemano, por ejemplo, diferencia cuadrada, diferencia absoluta, entre otras. Se ha discutido sobre cuál criterio debería utilizarse, ya que afecta en el resultado final, Qin et al., 2013 utilizó la distancia euclidiana, Hotaling et al., 2010c señala un método con un nuevo parámetro  $w$  que considera la distancia entre alternativas menos competitivas, finalmente Berkowitsch et al., 2015 creó una función generalizada para tres o más atributos.

### 2.5.1.2.1 Ejemplo de la formación de valores de preferencia

En los siguientes párrafos se ejemplifican los mecanismos básicos del modelo. La información para utilizar estos ejemplos son sacadas de la charla realizada por Hess (2018), la cual fue dictada en el Interdisciplinary Choice Workshop realizado en la ciudad de Santiago de Chile en Agosto de 2018<sup>1</sup>. Esta charla no está documentada, sin embargo, el ejemplo puede desarrollarse sin especificar la fuente, ya que corresponde solamente a calcular las matrices de DFT utilizando los valores del ejemplo. Se ejemplifica mediante un caso de elección de 4 vinos, que tienen los siguientes atributos observables: nombre, año, tipo, país de origen y precio. Los valores de estos son encontrados en la Tabla 2.2.

Como se mencionó, para el cálculo de la preferencia presente  $P_t$ , se debe multiplicar la preferencia pasada  $P_{t-1}$  por la matriz de retroalimentación  $S$ , y sumarle a esta multiplicación el valor presente entregado debido a la observación de atributos  $V_t$  (Ecuación 2.8).

Para el cálculo de  $V_t$  se necesita la matriz  $M$  (Tabla 2.2) que contiene el valor de los atributos para cada alternativa. Ésta debe ser multiplicada por la matriz de contraste (Tabla 2.3), que centra los valores alrededor de 0, y también multiplicada por el vector  $W_t$ , que indica cuál atributo fue observado por el individuo en el instante  $t$  (Ecuación 2.9). Por ejemplo, si se observa el atributo precio en el instante  $t$ , el vector  $W_t$  es igual a  $(0, 0, 0, 0, 1)$ .

M Matriz de Atributos	nombre	tipo	país	año	precio
Hess Select	3	3	3	1	-3
Campo Viejo	2	1	1	3	-2
Santa Rita	1	3	2	1	-2
Tesco Australian	0	2	0	0	-1

Tabla 2.2: DFT: Ejemplo de matriz de atributos  $M$

De la Tabla 2.2, se puede observar 2 características importantes, las cuales son consideradas como parte de las limitaciones del modelo. La primera característica, es que se puede notar que los efectos de las variables ya están definidos, es decir, se asume de antemano que las primeras 4 variables tienen efecto positivo en las preferencias de la persona. En cambio, la última variable (costo) se considera con efecto negativo. Esto ilustra una de las posibles limitaciones del modelo, en cuanto a que no es claro qué hacer con aquellas variables para las cuales no se conozca la dirección de sus preferencias a priori.

<sup>1</sup>Link ICW: <http://www.icw.cl/11/default/inicio> (Fecha revisión 19.09.2019)

Un segundo ajuste que es necesario considerar para aplicar el modelo DFT en la práctica tiene que ver con que los valores de las variables están definidos como valores de categoría 0, 1, 2 o 3, lo que supone que los valores de las variables originales fueron modificados para poder ajustarse al modelo DFT. El modelo no utiliza los valores reales de las variables, ya que, de ser así, el vector  $D^2$  tendría valores extremos y muy diferentes en valor y magnitud entre sí<sup>2</sup>. Luego, para obtener el valor de  $V_t$  se necesita del vector  $W_t$  y la matriz  $C$  (Tabla 2.3).

Matriz C	Hess Select	Campo Viejo	Santa Rita	Tesco Australian
Hess Select	1	-1/3	-1/3	-1/3
Campo Viejo	-1/3	1	-1/3	-1/3
Santa Rita	-1/3	-1/3	1	-1/3
Tesco Australian	-1/3	-1/3	-1/3	1

Tabla 2.3: DFT: Ejemplo de matriz de contraste C

El atributo que se observe depende de la probabilidad de que éste sea observado, por lo tanto, el vector  $w$  define estas probabilidades para todos los atributos, las cuales deben ser estimadas.

Definido todo lo anterior, y considerando el error  $\varepsilon$ , que debe ser estimado, se obtiene el valor del vector de valencia  $V_t$ . Luego, se debe obtener el valor de la matriz de retroalimentación  $S$  para tener todo lo necesario para el cálculo de  $P_t$ . Como se definió en la Ecuación 2.10,  $S$  depende de  $\phi_2$ ,  $\phi_1$  y  $D^2$ . Los dos primeros parámetros deben ser estimados y la matriz  $D^2$  se calcula de acuerdo a la diferencia cuadrada de los valores de los atributos. Por ejemplo, el valor de  $d_{Hess,Campo}$  que se observa en la Tabla 2.4 se calcula de la siguiente manera:  $d_{Hess,Campo} = (3 - 2)^2 + (3 - 1)^2 + (3 - 1)^2 + (1 - 3)^2 + (-3 - (-2))^2 = 14$

Matriz $D^2$	Hess	Campo	Santa Rita	Tesco
Hess	0	14	6	24
Campo	14	0	10	16
Santa Rita	6	10	0	8
Tesco	24	16	8	0

Tabla 2.4: DFT: Ejemplo de matriz de distancia D

<sup>2</sup>Durante el trabajo de esta memoria se intentó replicar el modelo DFT mediante codificación propia, con valores originales de variables (no de categoría) pero ocurría lo mencionado, es decir, la matriz  $D^2$  se no permitía un cálculo correcto del modelo.

Como se mencionó, los criterios que se elijan para el cálculo de la matriz  $D$  dependen del modelador. Por lo que no se concluye con exactitud cuál es la forma correcta de calcular esta matriz.

Finalmente, siguiendo la formula, se puede obtener la matriz  $S$  (Tabla 2.5) asumiendo que  $\phi_1 = 0,05$  y  $\phi_2 = 0,1$ , para el ejemplo.

Matriz $S$	Hess	Campo	Santa	Tesco
Hess	0.900	-0.050	-0.074	-0.030
Campo	-0.050	0.900	-0.061	-0.045
Santa	-0.074	-0.061	0.900	-0.067
Tesco	-0.030	-0.045	-0.067	0.900

Tabla 2.5: DFT: Ejemplo de matriz de retroalimentación  $S$

Con todos estos valores obtenidos, se tiene casi todo lo necesario para poder calcular los vectores de preferencias en cada instante  $t$ , lo único que falta en este ejemplo es el punto de partida  $P_0$ . Este punto de partida puede ser considerado como la opinión pre-análisis de las alternativas, que puede considerar también el valor de atributos que no se consideran en la modelación. Estos valores claramente no son conocidos, por lo que deben ser estimados.

### 2.5.1.3. Estimación

Para la estimación, se requiere conocer el valor esperado de  $P_t$  y la covarianza de  $P_t$ , ya que la probabilidad de elegir una alternativa sobre otra depende de estos valores. Roe et al. (2001) demuestra que  $P_t$  converge a una distribución normal multivariable. Entonces la expresión para el cálculo de la probabilidad de elegir la alternativa A, entre A, B y C, se ve en la Ecuación 2.11.

$$\mathbb{P}(A | t) = \mathbb{P}(P_t[A] - P_t[B] > 0 \cap P_t[A] - P_t[C] > 0) = \int_{X>0} \frac{e^{-(X-\Gamma)'\Lambda^{-1}(X-\Gamma)/2}}{2 \cdot \pi |\Lambda|^{0,5}} dX \quad (2.11)$$

donde

$$X = (P_t[A] - P_t[B], P_t[A] - P_t[C])$$

$$\Gamma = L \times E[P_t]$$

$$\Lambda = L \times Cov[P_t] \times L'$$

$$L = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

$L$  es una matriz que tiene una columna con 1s y una matriz identidad de tamaño  $n - 1$  ( $n$  numero de alternativas), la columna de 1s debe ser puesta en la columna de su alternativa respectiva (en este caso la primera, es decir, la de A). El cálculo de la esperanza y la covarianza se mejoró por Hancock et al., 2018, debido a que antes solo podía calcularse para  $t$  infinito. Luego, las expresiones finales se obtienen en las ecuaciones

$$E[P_t] = (I - S)^{-1}(1 - S^t) \cdot C \cdot M \cdot w_m + S^t \cdot P_0 \quad (2.12)$$

$$Cov[P_t] = (I - Z)^{-1}(I - Z^t)\bar{\Phi} \quad (2.13)$$

Con  $w_m$  vector de los pesos de los atributos,  $Z$  una transformación de  $S$  ( $\overline{S\Phi S'} = Z\bar{\Phi}$ ).

Se puede apreciar que esta formulación de la probabilidad de elección considera un gran uso de matrices que se operan matemáticamente entre ellas. Además, se debe tener en cuenta que lo anterior es solamente considerado para un individuo, por lo tanto, de aplicarse a toda una muestra de individuos, se vuelve aún más complejo el problema de computación.

#### 2.5.1.4. Analogía con MNL

Hancock et al. (2018) sugiere una interpretación aproximada para entender los mecanismos de DFT. Los atributos de las alternativas  $M$  son multiplicados por  $W$ , la importancia relativa de los atributos, que son equivalentes a los coeficientes  $\beta$  de MNL. Luego se tiene el vector de valencia  $V_t$ , que puede ser considerado como “la utilidad en un momento específico” y  $P_t$  la preferencia total, equivalente a la utilidad de las alternativas de MNL. DFT no produce utilidades, pero se puede usar el total de preferencias para calcular la verosimilitud de las alternativas.

### 2.5.1.5. Diferencia con otros modelos de la literatura

En la literatura se encuentran diversos estudios y propuestas de modelos que apuntan en la misma dirección que el modelo Decision Field Theory, es decir, que buscan considerar el proceso cognitivo y las estrategias que toman los individuos, profundizando el conocimiento en la dinamicidad de las decisiones. En los siguientes párrafos se pretende dar una noción de estos modelos y sus diferencias con DFT. Se debe aclarar, que sólo se dará una noción general de estos modelos, ya que estos solamente sirve como forma de comparación.

Los modelos lexicográficos (Slovic, 1975) son modelos no compensatorios que suponen que los individuos siguen ciertas reglas de ordenamiento para realizar la elección de alternativas. Por ejemplo, un ordenamiento lexicográfico de importancia de atributos supone que los individuos observan un atributo de todas la alternativas y ordenan las alternativas de mejor a peor, luego, observan otro atributo, y se repite el proceso hasta tener un ordenamiento de todos los atributos y sus alternativas. Finalmente el individuo escogería una estrategia para elegir la alternativa, por ejemplo, elegir la alternativa que mejor está rankeada en los atributos, o solamente elegir la alternativa que mejor rankea en el atributo más importante, etc. Se señala solamente el modelo lexicográfico por atributos, porque es el que mayor relación tiene con el modelo a estudiar en esta memoria, ya que también asume que los individuos realizan búsqueda en anchura (breadth-first). Luego, Jedidi et al. (2015), proponen incorporar incertidumbre en el orden de preferencia de los atributos, dejándolo como un modelo estocástico. Este modelo, a pesar de no ser publicado, se encuentra explicado por Henriquez (2014).

Otro modelo que entra en la discusión es el Elimination by Aspects Model (Tversky, 1972), el cual señala que cuando un individuo se enfrenta a diferentes opciones, primero identifica los atributos más importantes, luego, si alguna alternativa no cumple con criterios individuales esta es eliminada de ser una opción a escoger. Por ejemplo, al realizar un análisis de qué opción seleccionar, si una alternativa es extremadamente cara en comparación a las otras, algunos individuos tienden a eliminar esa alternativa completamente de sus opciones. Es claro porqué este modelo es diferente a DFT, sin embargo, este también tiene cercanía teórica ya que asume que los individuos utilizan una estrategia observación de atributo por atributo.

Dentro de la literatura, se encuentran también los modelos de búsqueda, que siguen la misma lógica del principio DFT, en donde se asume que el individuo realiza una observación continua y dinámica de los atributos para realizar la elección. Estos modelos se enfocan en la adquisición y procesamiento de la información, es decir, en las estrategias que éstos utilizan para observar e interpretar ésta, dependiente de la estrategia que utilicen. En general se calibran mediante la utilización de información de tecnologías como eye-tracking. Entre las propuestas más representativas de los modelos de búsqueda se encuentran los modelos

Information Acquisition Model (Shi et al., 2013), Search and Choice Model (Yang et al., 2015) y Satisficing Choice Model (Stüttgen et al., 2012). Information Acquisition Model es un modelo jerárquico de Markov de diferentes capas que utiliza información de experimentos de eye-tracking. En principio, este asume que existen 3 niveles fundamentales en el análisis de decisiones, el primero es el nivel del movimiento de ojos, luego, el cómo los individuos adquieren y procesan la información observada, y finalmente, la última capa se refiere a la estrategia que se utiliza para analizar, por ejemplo, comparando atributos por atributo. Search Choice Model es un modelo que también se basa en la utilización de eye-tracking, mediante el cual utilizan una función de utilidad de cada caso de observación, en el cual estiman parámetros a nivel individual y una función likelihood basada en métodos bayesianos. Si bien, estos 2 modelos y DFT tienen una naturaleza similar de buscar explicar el proceso cognitivo y las estrategias que se utilizan, la mayor diferencia radica en que estos modelos de búsqueda dependen de la estimación y predicción mediante utilización de tecnologías de eye-tracking, en cambio, DFT busca ser una alternativa modelos tradicionales que utilizan bases de datos sin esta información. Sin embargo, el eye-tracking puede ser utilizado para DFT con el fin de mejorar la precisión de éste. Satisficing Choice Model es un modelo que también sigue la lógica detrás de estos modelos, ocupa un concepto de satisfacción relacionada a los atributos y variables. Este concepto es muy similar al del modelo de Elimination by Aspects, en el cual si una alternativa no satisface los criterios del individuo, termina siendo descartada. Es clave entender que estos modelos se diferencian en la forma de calibración, ya que requieren la información de lo observado por los individuos, en cambio DFT, puede utilizar esta información, pero para complementar, lo mismo aplica para el modelo desarrollado en esta memoria.

#### **2.5.1.6. Limitaciones del modelo**

Las limitaciones y/o problemas del modelo que se mencionarán se dividirán en dos categorías, la primera es la que señalan Hancock et al. (2018), autores de la publicación más reciente que estudia el modelo, y la segunda es la que se identificaron a partir del trabajo de esta memoria.

Hancock et al. (2018) declaran las siguientes limitaciones:

1. Alto costo computacional: Un modelo estándar DFT puede tomar hasta 200 veces más que un Logit multinomial.
2. Interpretabilidad económica: Debido a la naturaleza de la formulación del modelo DFT no se puede realizar interpretaciones del valor de los parámetros o sus relaciones, por ejemplo, no se puede obtener el valor del tiempo.

3. Identificación: Signo parámetro de atributos: Es claro que existen parámetros que tienen efectos negativos y positivos para los individuos. Sin embargo, en DFT, siempre debe ser el parámetro  $w_i$  un valor positivo, ya que es una probabilidad. Por lo tanto, para los parámetros que son negativos, el modelo auto-fija su valor de probabilidad en 0, lo que no es realista, ya que significa que la probabilidad de observar un parámetro es menor, solo por el hecho de ser negativo. Esto se modifica cambiando el signo de los datos que tienen efecto negativo.
4. Configuración puntos iniciales: Se declara que DFT no tiene puntos iniciales claros para su estimación, por lo que se debe realizar ensayos para encontrar puntos iniciales adecuados. Por ejemplo, el autor señala que se realizan 100 simulaciones para encontrar puntos iniciales adecuados en su investigación.

Se probó esta hipótesis dentro del trabajo de esta memoria, ya que se intentó configurar el modelo con el software Apollo (Hess y Palma, 2019a), y, en efecto, al cambiar los puntos iniciales éste entregaba diferentes resultados, lo que sugiere que el modelo es no convexo. Lo anterior es una gran limitante, ya que se debe hacer pruebas de ensayo y error para poder obtener resultados con este modelo.

5. Escala de valores de atributos: La escala que se utiliza para los valores de los atributos de las alternativas cambia el resultado final de la estimación de un modelo DFT.

Luego, las identificadas para el trabajo de esta memoria son:

6. Alta dificultad de configuración: Se identificó que el modelo tiene una alta complejidad para poder configurarlo y poder realizar estimaciones. Esto debido a su a su cantidad de matrices que se deben crear, incluida su estimación, que crea transformaciones de matrices y incluir su distribución normal multivariada.
7. Signo de atributos: Como se pudo observar en el ejemplo entregado, se debe saber de antemano cuál es el efecto del atributo que se está estudiando. Por lo tanto, de no saberse con exactitud si un atributo tiene un efecto positivo o negativo, el modelo puede estar calculándose incorrectamente.
8. Alternativas con mismos atributos: Se descubrió, que al estimar el modelo con atributos que tienen el mismo valor para diferentes alternativas, causa que el modelo DFT no pueda ser estimado. Esto debido a que, según lo observado y según la formulación de Hancock et al. (2018), DFT depende de estimar matrices inversas, las cuales no pueden ser calculadas si dos o más alternativas tienen los mismos valores en algún atributo.



Existen dos posibles soluciones para lo anterior, lo cual limita aún más el modelo. La primera, es fijar los parámetros  $\phi_2$  y  $\phi_1$  en 0, ya que de esta forma no es necesario invertir matrices, lo cual entrega una versión simplificada de DFT, ya que se descubrió que estos parámetros son insignificantes a veces<sup>3</sup>. Y la segunda opción es eliminar del análisis los datos en que los atributos de las alternativas son iguales, lo que es una gran limitante, ya que se pierde poder explicativo, y además, no es posible compararlo con otros modelos que utilizan la base de datos completa.

---

<sup>3</sup>Thomas Hancock, comunicación personal via e-mail. Respuesta pública en foro de usuarios de software Apollo: <https://groups.google.com/d/msg/apollo-choice-modelling/ATi4cFWgvr4/ZkSyx9IOBQAJ>. 13 de Septiembre 2019.

## Capítulo 3

# Formulación de un modelo DFT basado en el principio de Maximización de la Utilidad Aleatoria

Al igual que DFT, el modelo propuesto está ideado principalmente para contextos de elecciones de corto plazo, es decir, elecciones que no toman más de unos segundos o minutos de análisis, esto debido a que según Hancock et al. (2018) aún no se tiene claridad respecto a qué situaciones DFT puede aplicarse con seguridad el modelo DFT, pero se ha probado el modelo con bases de datos que contienen situaciones de elecciones de corto plazo (Encuestas PD y PR).

Este modelo asume que los individuos tienen una cantidad limitada de alternativas con sus respectivos atributos, y que, para tomar una decisión respecto a qué alternativa escoger, realizan una evaluación secuencial de los atributos. En cada paso de este proceso, el individuo observa un atributo de todas las alternativas, se genera una opinión de éstas opciones y luego sigue buscando en base a los atributos que tiene disponible. Si es que le parece que se formó una opinión sólida de las alternativas, y cree que su opinión no cambiará a futuro, la persona escoge finalmente una de las alternativas.

En este capítulo se presenta la base general y matemática del modelo propuesto. Es importante destacar que lo presentado a continuación y durante toda la memoria corresponde un prototipo de modelo, que busca hacerse cargo de algunas de las limitaciones del modelo Decision Field Theory. Por lo tanto, primero se formulará el modelo, con todo lo necesario para que pueda desarrollarse un método de generación de datos y de estimación, pero luego en el Capítulo 4 y 5 se analizará la validez y todos los casos de posibles ajustes que se

necesiten en futuras iteraciones, sentando así las bases para un trabajo futuro.

### 3.1. Ejemplo introductorio

En esta sección se realizará un pequeño ejemplo para poder relacionar la teoría con un caso que podría ocurrir antes de realizar una elección. Es importante destacar que se intenta replicar lo que puede ocurrir realmente.

Se supone el caso de un individuo que quiere viajar a otra ciudad de su país. Este individuo está totalmente decidido a elegir pronto, es impaciente y las únicas opciones que baraja son 3: avión, bus y tren. Los atributos visibles de estas alternativas son el precio del pasaje, el tiempo de viaje aproximado y la hora de salida. Antes de observar las opciones, el individuo prefiere el modo tren, dada su experiencia e historia, y sabe que será una opción más segura. Sin embargo, está dispuesto a analizar otras opciones.

La información la tiene disponible al mismo tiempo, y siguiendo el supuesto que se ha mencionado en esta memoria, el individuo realiza una observación atributo por atributo.

Siguiendo los supuestos del modelo, el individuo inicia su proceso, observa un atributo, actualiza su utilidad, observa otro atributo y nuevamente actualiza su utilidad, esto lo realiza hasta que quede conforme con su búsqueda y termina el proceso, eligiendo una alternativa.

### 3.2. Modelo econométrico

La diferencia general del modelo propuesto en esta investigación con Decision Field Theory es que este modelo tiene otra formulación matemática y realiza otros supuestos. Como se mencionó, este modelo busca considerar en su formulación la evaluación de atributos en el tiempo. Se asume, entonces, que existe una utilidad subyacente que cambia a medida que se observan los atributos. La decisión de seguir observando o no depende de un umbral de tolerancia que se vuelva más restrictivo mientras avanza el tiempo. Además, se asumirá que los atributos pueden ser visitados más de una vez, y se pueden visitar consecutivamente, emulando lo que considera el método DFT.

Los supuestos fundamentales del modelo propuesto son:

- Se asume que los individuos realizan observación búsqueda en anchura (breadth-first search), es decir, observan atributo por atributo en cada paso, al igual que Decision Field Theory (Busemeyer y Townsend, 1993; Hancock et al., 2018; Roe et al., 2001).

Sin embargo las alternativas se comparan solamente al momento de elegir (al menos explícitamente), a diferencia de Decision Field Theory, que asume que los individuos sí realizan observación por atributo pero que comparan las alternativas en cada instante de observación.

- Por lo tanto, el número de pasos es ilimitado, pero para efectos de esta memoria, es decir, para la generación de datos y para la estimación los pasos se limitan de acuerdo a un número de pasos máximos a realizar (tiempo máximo) que resulta factible para el análisis, tal como ocurre en la realidad en casos de elecciones en contextos de corto plazo. Además, es lógico pensar que los pasos son limitados, ya que o sino, los individuos nunca escogerían.
- Además, al intentar replicar el comportamiento DFT, los atributos observados pueden ser observados más de una vez, y observados consecutivamente. Lo anterior se debe a que Busemeyer y Townsend (1993) señala que los pasos son periodos de tiempo de análisis cognitivo que el individuo realiza, por lo tanto, intenta replicar el tiempo que un individuo realiza su proceso de deliberación de forma lineal. No se tiene claridad definida de cuanto tiempo (en segundos) un paso de análisis debe durar, pero por ejemplo Busemeyer y Townsend (1993) señala que estos tienen a 0, para poder realizar un análisis continuo. Sin embargo, para este modelo prototipo, y con el fin de buscar simplicidad, se asume que el tiempo total es igual a la cantidad de pasos, es decir, se considerará que cada paso tiene una duración de un segundo (o unidad unitaria).

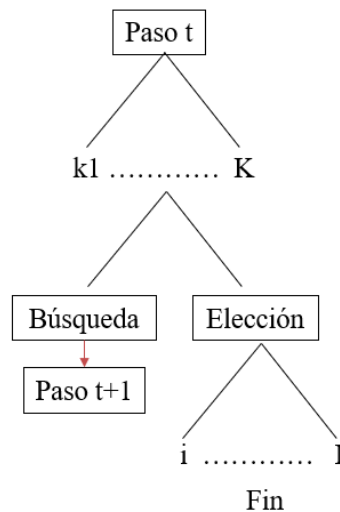


Figura 3.1: Decisiones involucradas en cada paso  $t$

Se define el conjunto de alternativas a elegir como  $I$  y el conjunto de atributos como  $K$ . La utilidad de la alternativa  $i$  del individuo  $n \in N$  en cada paso  $t \in T$  es  $U_{ni}^t$ . En  $t = 0$  el individuo tiene una opinión inicial preconcebida de cada alternativa (que el investigador

desconoce) definida como  $U_{ni}^0$ . Luego, en cada paso  $t$ , tiene dos opciones a tomar: decidir elegir, o seguir observando atributos y actualizar su utilidad (Figura 3.1). Si se decide a elegir, se termina el proceso, la utilidad deja de actualizarse y se procede a escoger la que tenga mayor utilidad de las alternativas  $I$ . En cambio, si la persona decide seguir observando atributos, debe determinar cuál de éstos observar para continuar actualizando su utilidad (cambiar su opinión o reforzarla).

La utilidad de cada alternativa se actualiza de acuerdo al atributo  $k$  que se observó en ese paso  $t$ . El proceso se repite hasta que la persona elige, por lo que éste puede ser interrumpido en cualquier paso  $t$ . Si el proceso llega al tiempo máximo límite, el individuo está forzado a elegir en ese paso.

Consideremos que la persona elige la alternativa  $i$  siguiendo una secuencia específica. Por ejemplo, supóngase que la persona observó atributo  $k_3$  en  $t = 1$ , luego observó  $k_2$  en  $t = 2$ , después el mismo atributo  $k_2$  en el tercer momento de observación  $t = 3$  y finalmente realizó su elección, la secuencia se denota como  $(k_3, k_2, k_2, i)$ .

Entonces, la probabilidad de dicho suceso  $P(k_3, k_2, k_2, i)$  se define como la probabilidad de observar esos atributos en sus respectivos momentos y la probabilidad de elegir  $i$  dado que se observaron esos atributos en ese orden, como se puede ver en la Ecuación 3.1. Luego, utilizando el teorema de Bayes, se puede escribir una expresión, tal y como se ve en la Ecuación 3.2, para cualquier combinación de atributos del conjunto  $K$  observados en una cantidad  $T$  de pasos. Finalmente, una secuencia de observación de atributos se define como caminos de búsqueda o, simplemente, caminos.

Entonces, para el caso mencionado y asumiendo independencia entre la probabilidad de elegir atributo  $k$ :

$$\begin{aligned} \mathbb{P}(k_3, k_2, k_2, i) &= \mathbb{P}(i \mid k_3, k_2, k_2) \cdot \mathbb{P}(k_3, k_2, k_2) \\ \mathbb{P}(k_3, k_2, k_2, i) &= \mathbb{P}(i \mid k_3, k_2, k_2) \cdot \mathbb{P}(k_3 \mid t = 1) \cdot \mathbb{P}(k_2 \mid t = 2) \cdot \mathbb{P}(k_2 \mid t = 3) \end{aligned} \quad (3.1)$$

$$\mathbb{P}(k(1), \dots, k(t), i) = \mathbb{P}(i \mid k(1), \dots, k(t)) \prod_{\tau=1}^T \mathbb{P}(k(\tau)) \quad (3.2)$$

### 3.2.1. Toma de decisión en cada paso $t$

¿Cómo el individuo toma decisiones en cada paso  $t$ ? Como se mencionó y definió anteriormente, después de haber actualizado su utilidad en el paso  $t$ , existen 2 opciones principales: se debe decidir si seguir observando (continuar proceso) o si elegir (terminar proceso). Si se

decide continuar con la búsqueda ( $s(t)$ ), se debe determinar qué atributo  $k$  se quiere observar para actualizar la utilidad en  $t + 1$ . Si se decide elegir ( $e(t)$ ), se debe decidir qué alternativa elegir en  $t$ .

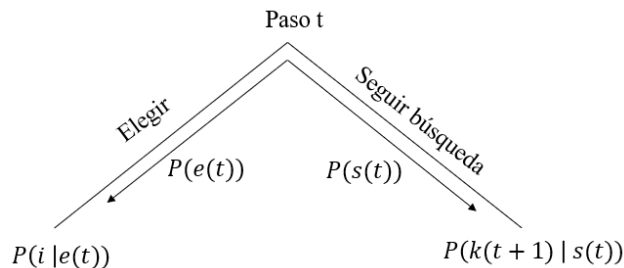


Figura 3.2: Probabilidades asociadas en cada paso  $t$

Con el fin de simplificar la notación de combinaciones de atributos ( $k(1)...(k(T))$ ) se define el conjunto  $H^t$ , como el conjunto de todos los caminos posibles que puede tomar el individuo hasta el momento  $t$ , es decir, este conjunto contiene todas las combinaciones existentes de atributos pertenecientes a  $K$  hasta  $t$ . Entonces, se define un camino cualquiera hasta  $t$  como  $h_i^t$ , con  $l$  siendo un número cualquiera de 1 hasta el total  $L$  de combinaciones existentes hasta el instante  $t$  ( $L = \sum_{\tau}^T |I|^{\tau}$ , con  $|I|$  el total de alternativas). La probabilidad de elegir un atributo  $k$  en el paso  $t$  depende de si se decidió seguir observando o no en el paso anterior  $t - 1$  (Ecuación 3.3). Así mismo, la probabilidad de elegir la alternativa  $i$  depende de si se decidió elegir en  $t$  (Ecuación 3.4).

$$\mathbb{P}(k(t) | h_i^{t-1}) = P(k(t) | s(t-1)) \cdot P(s(t-1) | h_i^{t-1}) \quad (3.3)$$

$$\mathbb{P}(i | h_i^t) = P(i | e(t)) \cdot P(e(t) | h_i^{t-1}) \quad (3.4)$$

### 3.2.2. Función de utilidad

Definida la base probabilística general para representar las decisiones, se debe definir la función utilidad que considere el aspecto dinámico de la evolución de preferencias de los individuos. Se pretende definir una función que sea lo más sencilla posible y que tenga similitudes con la expresión de utilidad del modelo Multinomial Logit para una fácil interpretación, posible identificación y recuperación de parámetros. Por lo tanto, se define la utilidad subyacente del individuo  $n$  en cada paso  $t$  para cada alternativa  $i$  como  $U_{ni}^t$ . Esta expresión contiene un aspecto dinámico  $t$ , el cual depende del atributo  $k(t)$  que se observa en ese paso, además de la utilidad que se percibió en el instante anterior,  $t - 1$ . Luego, la expresión de utilidad que

se define para este modelo es:

$$U_{ni}^t = \alpha \cdot U_{ni}^{t-1} + (1 - \alpha) \cdot \beta_{k(t)} \cdot X_{nik(t)} + \varepsilon_{nik(t)} \quad (3.5)$$

Con

$U_{ni}^t$  : utilidad subyacente de la alternativa  $i$  para el individuo  $n$  en  $t$

$\alpha$  : peso del pasado, factor temporal

$k(t)$  : atributo observado en el paso  $t$

$\beta_{k(t)}$  : efecto del atributo  $k(t)$

$X_{nik(t)}$  : valor del atributo  $k(t)$  para el individuo  $n$  de la alternativa  $i$

$\varepsilon_{in}^t$  : error temporal, distribuye valor extremo con escala  $\mu_e$

Esta utilidad tiene una lógica similar a las cadenas de Markov (Markov, 1907), en donde solamente el evento anterior ( $U_{ni}^{t-1}$ ) afecta el evento presente. La utilidad depende de la utilidad que se tenía hasta el paso de observación anterior, y del atributo  $k(t)$  que se observa en ese instante  $t$ . Este valor del atributo multiplica a un parámetro  $\beta_{k(t)}$  que atrapa el efecto del atributo en la utilidad del individuo, al igual que MNL, pero su diferencia radica en que el parámetro tiene efecto en la utilidad sólo en los momentos de observación.

La utilidad anterior  $U_{ni}^{t-1}$  es multiplicada por el factor  $\alpha$  el cual se define como el peso del pasado. Se definió, que a pesar de que se busca modelar procesos de elección de corto plazo, aun así las observaciones más recientes tienen un efecto que se balancea con el pasado. Además, se considera el peso del presente, que es el peso del atributo observado en  $t$ ,  $k(t)$ , el cual se multiplica por el parámetro  $\beta_{k(t)}$ , y es ponderado por  $(1 - \alpha)$ .

Es importante destacar, que esta función tiene alta dependencia de la utilidad inicial (opinión inicial de cada alternativa)  $U_{ni}^0$ , ya que esta define la evolución que tomará la utilidad subyacente de los individuos, al menos en los primeros pasos  $t$ .

Por último, se descompone la utilidad  $U_{ni}^t$  en dos partes (Ecuación 3.6): la primera es la utilidad sistemática, que no tiene variabilidad estocástica y depende solamente de los valores exactos de los parámetros y las variables. La segunda parte es el error, el que define a la utilidad subyacente como aleatoria. Es necesario definir esto para el cálculo de las probabilidades que dependen de la distribución del error.

$$\underbrace{U_{ni}^t}_{\text{Utilidad aleatoria}} = \underbrace{V_{ni}^t}_{\text{Utilidad sistemática}} + \underbrace{\varepsilon_{ni}^t}_{\text{Error}} \quad (3.6)$$

Es necesario mencionar, que la forma de la utilidad siempre dependerá del camino que toma el individuo, por lo que la notación de esta función debiera incluir el orden de observación. Luego se define  $U_{ni}^{t|h_t}$  como la notación dependiente del camino, la cual, por simplificación, no será siempre utilizada, sino que cuando sea explícitamente necesaria. Por lo tanto, se debe asumir que  $U_{ni}^t$  es para cualquier camino de observación, si es que no se especifica.

### 3.2.3. Expresiones de las probabilidades

Ya definida la utilidad subyacente, la estructura de probabilidades y el proceso que se busca explicar, se pueden escribir las expresiones de probabilidades del modelo.

#### 3.2.3.1. Probabilidad de elegir o seguir observando

Para definir estas probabilidades se debe tener en consideración que el individuo en el paso  $t$ , está decidiendo si continuar con su búsqueda o no. Por lo tanto, éste debe considerar si existe alguna ganancia en seguir realizando su búsqueda. Para esto, se asume que el individuo realiza una comparación entre la utilidad presente de todas las alternativas y la utilidad posible futura en caso de que elija en el siguiente periodo. Es necesario aclarar, que la utilidad futura del individuo no es conocida por él, sin embargo, se utiliza el supuesto de que las probabilidades de observar a futuro son computaciones realizadas inconscientemente y la utilización de la utilidad futura es referencial al trade-off que un individuo podría realizar. En la Figura 3.3 se observa que el individuo tiene asociados 2 nodos, uno presente para las utilidades actuales y uno futuro para las utilidades posibles futuras en caso de seguir buscando.

Luego, la utilidad actual (presente) en  $t$   $V_{presente}^t$  de cada individuo se define como la esperanza del máximo de las utilidades presentes de cada alternativa en  $t$ , esto es similar a la expresión de Nested Logit (Ben-Akiva et al., 1985), realizando la analogía de que todas las utilidades en  $t$  de cada alternativa son parte de un nido.

Para obtener esta forma cerrada de la utilidad máxima esperada (Ecuación 3.7), se asume que  $\varepsilon_{ni}^t$  es variable aleatoria independiente e idénticamente distribuida (i.i.d.) y que distribuye



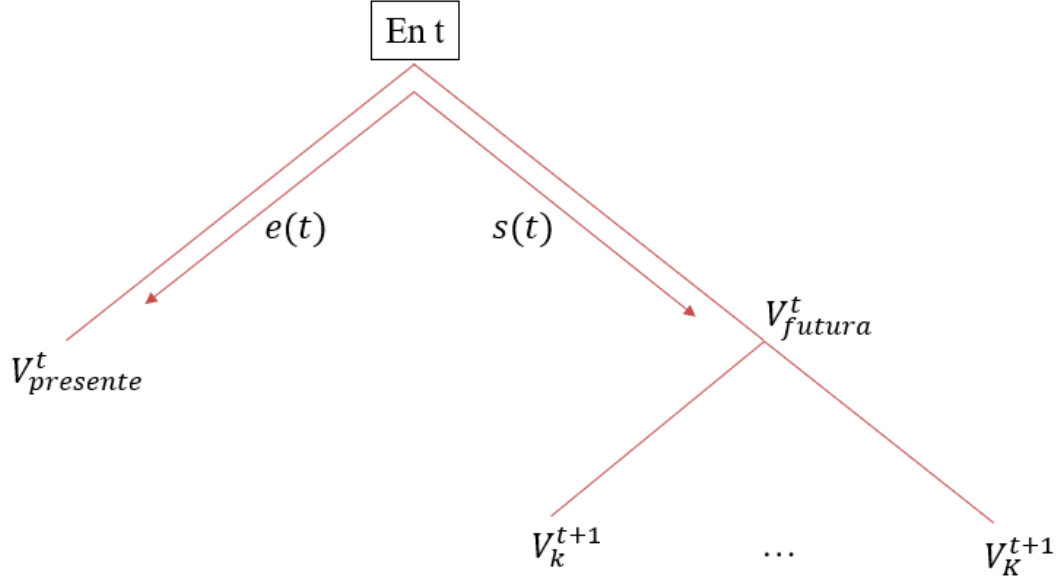


Figura 3.3: Utilidades asociadas a la decisión de seguir buscando o no valor extremo con escala  $\mu_e$  (Gumbel).

$$\begin{aligned}
 U_{presente}^t &= V_{presente}^t + \varepsilon_{presente}^t \\
 &\Downarrow \\
 V_{presente}^t &= E[\max(U_i^t \forall i \in I, \mu_e)] \\
 V_{presente}^t &= \frac{1}{\mu_e} \cdot \ln \left( \sum_i^I e^{\mu_e \cdot V_i^t} \right)
 \end{aligned} \tag{3.7}$$

Luego, la utilidad de seguir buscando (utilidad futura) se divide en 2 partes. Esto debido a que primero, se debe analizar la utilidad que podría entregar cada atributo  $k$  para todas las alternativas si se observase en  $t + 1$  ( $V_{sk}^{t+1}$ ), y luego, analizar esto para todos los atributos ( $V_{futura}^t$ ), como se puede observar en la Figura 3.3 y en las Ecuaciones 3.8 y 3.9.

Entonces, se define la utilidad futura posible de seguir buscando ( $s$ ) y observar el atributo  $k$ , como:

$$\begin{aligned}
 U_{sk}^{t+1} &= V_{sk}^{t+1} + \varepsilon_k^{t+1} \\
 &\Downarrow \\
 V_{sk}^{t+1} &= E[\max(U_{i|k}^t \forall i \in I, \mu_e)] \\
 V_{sk}^{t+1} &= \frac{1}{\mu_e} \cdot \ln \left( \sum_i^I e^{\mu_e \cdot V_{i|k}^{t+1}} \right)
 \end{aligned} \tag{3.8}$$

Teniendo esta expresión, se puede definir la utilidad futura posible, en caso de seguir buscando, como la esperanza del máximo de estos atributos, al igual que con la utilidad presente que tiene una estructura similar a un nido de Nested Logit. Para lograr esto se asume que  $\varepsilon_k^t$  es i.i.d valor extremo con escala  $\mu_s$ .

$$\begin{aligned}
U_{futura}^t &= V_{futura}^t + \varepsilon_{futura}^t \\
&\Downarrow \\
V_{futura}^t &= E[\max(U_{sk}^{t+1} \forall k \in K, \mu_s)] \\
V_{futura}^t &= \frac{1}{\mu_s} \cdot \ln \left( \sum_k^K e^{\mu_s \cdot V_{sk}^{t+1}} \right)
\end{aligned} \tag{3.9}$$

La probabilidad de seguir observando atributos para actualizar la utilidad depende de que la utilidad esperada futura supere un umbral de tolerancia  $\delta_t$  (3.10), que indica que existe una ganancia en continuar con el proceso para el individuo. Este factor de tolerancia crece a medida que pasa el tiempo, ya que, mientras más observa la persona, más impaciente se ha vuelto.

$$\delta_t = \delta \cdot t^2 \tag{3.10}$$

Luego, para que la persona continúe buscando, el valor absoluto de la diferencia entre la utilidad presente y la futura debe ser mayor a este umbral  $\delta_t$ . La intuición detrás de esto es que la persona solamente seguirá observando, si encuentra que a futuro existirán grandes cambios en su utilidad, pero a medida que pasa el tiempo, más intolerante se vuelve la persona, por lo que requiere cambios esperados mucho más grandes para decidir seguir buscando. De lo anterior sale la dinamicidad del parámetro  $\delta_t$ , como se muestra en la Ecuación 3.10. La formulación anterior no es definitiva, ya que se podría definir un factor constante (estable) base para  $\delta_t$  como  $\delta_0$ , con  $\delta_t = \delta_0 + \delta \cdot t^2$ , donde se podrían capturar dos posibles efectos, un efecto estable de los individuos y uno que aumenta en el tiempo, ambos deberían ser estimados. Pero para efectos de esta investigación exploratoria, se mantendrá simplicidad.

Dicho lo anterior y asumiendo que el error de la diferencia entre estas utilidades es i.i.d valor extremo con factor de escala  $\mu$ , se define:

$$\begin{aligned}
\mathbb{P}(s(t) | h_l^t) &= P(|V_{presente}^t - V_{futura}^t| \leq \delta_t) \\
\mathbb{P}(s(t) | h_l^t) &= \frac{e^{\mu \cdot V_{presente}^t}}{e^{\mu \cdot V_{presente}^t} + e^{\mu \cdot (V_{futura}^t + \delta_t)}} + \frac{e^{\mu \cdot V_{futura}^t}}{e^{\mu \cdot V_{futura}^t} + e^{\mu \cdot (V_{presente}^t + \delta_t)}}
\end{aligned} \tag{3.11}$$

Y la probabilidad de decidir elegir en  $t$  como :

$$\mathbb{P}(e(t) | h_i^t) = \frac{e^{\mu \cdot (V_{futura}^t + \delta_t)}}{e^{\mu \cdot V_{presente}^t} + e^{\mu \cdot (V_{futura}^t + \delta_t)}} + \frac{e^{\mu \cdot (V_{presente}^t + \delta_t)}}{e^{\mu \cdot V_{futura}^t} + e^{\mu \cdot (V_{presente}^t + \delta_t)}} \quad (3.12)$$

### 3.2.3.2. Probabilidad de observar atributo $k$

Se define la probabilidad de observar un atributo como el peso de cada utilidad futura de cada atributo ( $V_{sk(t+1)}^{t+1}$ ) entre todos los atributos posibles a considerar, asumiendo que  $\varepsilon_k^t$  es i.i.d valor extremo con escala  $\mu_s$  por lo tanto:

$$\mathbb{P}(k(t+1)|s(t)) = \frac{e^{\mu_s \cdot V_{sk(t+1)}^{t+1}}}{\sum_{k'}^K e^{\mu_s \cdot V_{sk'}^{t+1}}} \quad (3.13)$$

### 3.2.3.3. Probabilidad de elegir alternativa $i$

La probabilidad de elegir una alternativa es similar a la del modelo Logit, debido a la distribución del error  $\varepsilon_{ni}^t$  (valor extremo con escala  $\mu_e$ ).

$$\mathbb{P}(i(t)|e(t)) = \frac{e^{\mu_e \cdot V_i^t}}{\sum_j^I e^{\mu_e \cdot V_j^t}} \quad (3.14)$$

### 3.2.3.4. Expresiones finales

Finalmente, se tienen las expresiones finales de las ecuaciones (3.3) y (3.4):

$$\mathbb{P}(k(t) | h_i^{t-1}) = \frac{e^{\mu_s \cdot V_{sk(t)}^t}}{\sum_{k'}^K e^{\mu_s \cdot V_{sk'}^t}} \cdot \underbrace{\left( \frac{e^{\mu \cdot V_{presente}^{t-1}}}{e^{\mu \cdot V_{presente}^{t-1}} + e^{\mu \cdot (V_{futura}^{t-1} + \delta_t)}} + \frac{e^{\mu \cdot V_{futura}^{t-1}}}{e^{\mu \cdot V_{futura}^{t-1}} + e^{\mu \cdot (V_{presente}^{t-1} + \delta_t)}} \right)}_{\mathbb{P}(s(t-1) | h_i^{t-1})} \quad (3.15)$$

$$\mathbb{P}(i | h_i^t) = \frac{e^{\mu_e \cdot V_i^t}}{\sum_j^I e^{\mu_e \cdot V_j^t}} \cdot \underbrace{\left( \frac{e^{\mu \cdot (V_{futura}^t + \delta_t)}}{e^{\mu \cdot V_{presente}^t} + e^{\mu \cdot (V_{futura}^t + \delta_t)}} + \frac{e^{\mu \cdot (V_{presente}^t + \delta_t)}}{e^{\mu \cdot V_{futura}^t} + e^{\mu \cdot (V_{presente}^t + \delta_t)}} \right)}_{\mathbb{P}(e(t) | h_i^t)} \quad (3.16)$$

### 3.2.4. Ejemplo

Siguiendo con el ejemplo (3.1), supongamos que el individuo realizó el siguiente proceso: observó primero el tiempo de viaje ( $t_v$ ) de cada alternativa, luego, el precio ( $p$ ), y finalmente, observó el horario ( $h$ ) y se decidió por la opción tren.

Luego, el camino tomado por el individuo se denota  $(t, p, h, \text{tren})$ . Entonces, la probabilidad de elegir la opción tren y que se haya seguido ese camino se calcula como:

$$\mathbb{P}(t_v, p, h, \text{tren}) = \mathbb{P}(\text{tren} \mid t_v, p, h) \cdot \mathbb{P}(t_v \mid t = 1) \cdot \mathbb{P}(p \mid t_v) \cdot \mathbb{P}(h \mid t_v, p) \quad (3.17)$$

Con las probabilidades de observar los atributos dado que se decidió seguir observando:

$$\begin{aligned} \mathbb{P}(t_v \mid t = 1) &= \frac{e^{\mu_s \cdot V_{st_v}^1}}{e^{\mu_s \cdot V_{st_v}^1} + e^{\mu_s \cdot V_{sp}^1} + e^{\mu_s \cdot V_{sh}^1}} \cdot \mathbb{P}(s(t = 1)) \\ \mathbb{P}(p \mid t_v) &= \frac{e^{\mu_s \cdot V_p^{2|t_v}}}{e^{\mu_s \cdot V_{st_v}^{2|t_v}} + e^{\mu_s \cdot V_{sp}^{2|t_v}} + e^{\mu_s \cdot V_{sh}^{2|t_v}}} \cdot \mathbb{P}(s(t = 2) \mid t_v) \\ \mathbb{P}(h \mid t_v, p) &= \frac{e^{\mu_s \cdot V_h^{3|(t_v, p)}}}{e^{\mu_s \cdot V_{st_v}^{3|(t_v, p)}} + e^{\mu_s \cdot V_{sp}^{3|(t_v, p)}} + e^{\mu_s \cdot V_{sh}^{3|(t_v, p)}}} \cdot \mathbb{P}(s(t = 3) \mid t_v, p) \end{aligned} \quad (3.18)$$

Con  $V_{sk}^t(h_i^t)$  la utilidad del tiempo  $t$  dado el camino  $h_i^t$  observado.

Luego, se tiene que la probabilidad de elegir tren, es:

$$\mathbb{P}(\text{tren} \mid t_v, p, h) = \frac{e^{\mu_e \cdot V_{tren}^{3|(t_v, p, h)}}}{e^{\mu_e \cdot V_{tren}^{3|(t_v, p, h)}} + e^{\mu_e \cdot V_{bus}^{3|(t_v, p, h)}} + e^{\mu_e \cdot V_{avión}^{3|(t_v, p, h)}}} \cdot \mathbb{P}(e(t = 3) \mid t_v, p, h) \quad (3.19)$$

## 3.3. Probabilidades latentes

Las fórmulas anteriores son solamente para el cálculo de la probabilidad de solo un posible camino que el individuo pudo haber realizado. Por lo tanto, debe generalizarse la probabilidad para todos los caminos posibles que pudo haber realizado el individuo, ya que como se mencionó anteriormente, no se cuenta con información de lo observado por el individuo.

Luego, por cada paso  $t$  se aumenta el espacio posible de caminos  $H^{t-1}$  a  $H^t$  en  $K^t$ . Por ejemplo, si un individuo realizó 2 observaciones de un conjunto de 3 atributos, y como no se conoce lo que se observó, en  $t = 1$  existen 3 opciones de lo que pudo observar, y luego existen 3 opciones en  $t = 2$  para cada posible opción elegida del paso anterior. Luego, la cantidad de caminos posibles que puede tomar el individuo es  $3 + 9 = 12$ . En la Figura 3.4 se observa la generalización de lo mencionado.

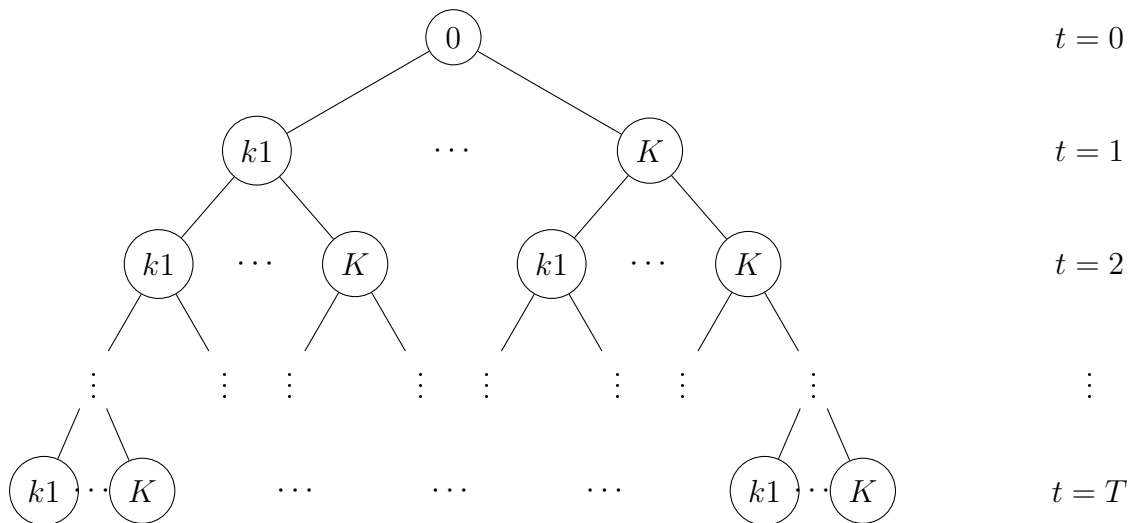


Figura 3.4: Generalización caminos posibles de un individuo

Sabiendo esto, se calcula la probabilidad de elegir una alternativa  $i$ , como la suma de la probabilidad de elegir  $i$  dado un camino por la probabilidad de elegir ese camino, para todos los caminos (Ecuación 3.20).

$$\mathbb{P}(i) = \sum_{h_i^t}^{H^T} \mathbb{P}(i | h_i^t) \cdot \mathbb{P}(h_i^t) \quad (3.20)$$

## 3.4. Estimación

### 3.4.1. Expresión matemática de la estimación

Teniendo una expresión para  $\mathbb{P}(i)$ , y definiendo la elección  $y_{ni}$  hecha por  $n$  de la alternativa  $i$  como:

$$y_{ni} = \begin{cases} 1 & \text{si individuo } n \text{ eligió } i \\ 0 & \text{si no } \sim \end{cases}$$

Se tiene que la verosimilitud de observar la información se escribe como:

$$L = \prod_n^N \prod_i^I \mathbb{P}(i)^{y_{ni}} \quad (3.21)$$

Finalmente, al aplicar logaritmo natural a esta expresión, se obtiene la log-verosimilitud, y por lo tanto, el modelo tiene todo lo necesario para poder estimarse, asumiendo que se tiene información de los atributos a estudiar y la elección realizada por los individuos.

$$LL = \sum_n^N \sum_i^I y_{ni} \cdot \ln(\mathbb{P}(i)) \quad (3.22)$$

La ecuación 3.22 asume que todos los puntos iniciales de cada alternativa, llamados opiniones iniciales,  $U_{ni}^0$  son iguales para todos los individuos, es decir  $U_{ni}^0 = U_i^0 \forall n \in N$ . Para efectos de esta memoria, se hará este supuesto, debido a la simplicidad que implica para probar el modelo. En trabajos futuros se deberá plantear la estimación con parámetros individuales de opinión inicial (o, incluso, para todos los parámetros) mediante la expresión de verosimilitud de la Ecuación 3.23. En donde  $\text{dist}(U^0 | \Omega)$  es la distribución de la utilidad inicial de la población.

$$L = \int_{U^0} \prod_n^N \prod_i^I \mathbb{P}(i|U^0)^{y_{ni}} \cdot \text{dist}(U^0 | \Omega) d\Omega \quad (3.23)$$

### 3.4.2. Métodos de estimación

En esta memoria, se realizó un análisis de prueba de 4 especificaciones para la estimación del modelo. Esto, debido a su complejidad inicial. Es importante destacar que las especificaciones 1, 2 y 4 serán las más estudiadas en el Capítulo 4. La especificación 3 fue descartada por su complejidad y poco rendimiento. Por último, la especificación 5 se explicará mas no se desarrollará en esta memoria, dejándola como una posible extensión para trabajos futuros

Las especificaciones son descritas a continuación:

### **3.4.2.1. Especificación 1: Estimación suponiendo que se conoce el camino de búsqueda realizado por el individuo**

Esta estimación supone que lo observado por el individuo en cada paso  $t$  es información conocida. Esta data puede obtenerse mediante nuevas tecnologías, por ejemplo, seguimiento del ojo de los individuos al realizar su proceso de decisión (Eye-tracking).

### **3.4.2.2. Especificación 2: Estimación suponiendo que se conoce la cantidad de pasos que le tomó al individuo**

Esta especificación considera que se conoce el tiempo o la cantidad de pasos que tomó el individuo. Esto se puede obtener mediante la implementación de herramientas que puedan capturar el tiempo de respuesta de los individuos, y luego, obtener cuánto tiempo se demoran los individuos en cada paso (para obtener la cantidad de pasos).

### **3.4.2.3. Especificación 3: Estimación asumiendo que la probabilidad de un camino es una clase latente**

Durante la investigación para esta memoria, se realizó una propuesta de calcular las probabilidades de cada camino de observación como independientes, es decir, que para camino  $h_i^t$ , existe una probabilidad asociada  $\mathbb{P}(h_i^t)$  única, es decir, esta probabilidad debería ser estimada. Lo anterior se asemeja al modelo Logit de clases latentes, en donde cada camino es una clase latente. Esto resultaba factible para casos en donde la cantidad de pasos era pequeña (1, 2 o 3) y con pocos atributos, ya que al agregar más atributos o cantidad de pasos el espacio de caminos se agranda considerablemente lo que obliga al cálculo de una gran cantidad de parámetros. Lo anterior no es óptimo, ya que el hecho de tener que estimar tantos parámetros complejiza el modelo considerablemente, por lo tanto, esta opción fue descartada.

### **3.4.2.4. Especificación 4: Estimación tradicional (probabilidades latentes)**

La mayoría de las bases de datos actuales de elecciones discretas contienen solamente información visible de los atributos, de los individuos y las elecciones realizadas por éstos, por lo tanto, es necesario contar con un modelo que pueda estimarse utilizando solo estos datos. Esta estimación es la desarrollada completamente en este capítulo.

### 3.4.2.5. Especificación 5: Estimación usando probabilidades definidas acortando el espacio de caminos posibles utilizando el enfoque propuesto por Guevara et al. (2018)

El gran problema de la estimación convencional, es que conlleva un cálculo matricial complejo, que se complejiza a medida que se agregan pasos de observación en la estimación. Por lo tanto, para trabajos futuros se propone implementar un método de estimación que acorte el espacio factible de caminos de búsqueda, sin grandes pérdidas en poder explicativo.

## 3.5. Codificación en R

Para poder representar, calcular y estimar este modelo, se utilizó solamente el lenguaje de programación R (R Core Team, 2017), y el software RStudio (RStudio Team, 2015), esto dado que se busca simplicidad y fácil implementación del modelo. Por lo tanto, agregar más softwares y lenguajes de programación complicaría este objetivo de simplificación. (Por ejemplo, MNL es de simple implementación en R)

En esta sección, se describirá de forma general el proceso de codificación para la estimación del modelo. El código general para la generación de datos y estimación se encuentra en la sección Anexos (Apéndice A y B).

Como se menciona en la sección 3.4, para poder estimar el modelo se necesita calcular y optimizar la función de verosimilitud. Para esto, primero se necesita saber la probabilidad de todos los caminos que el individuo puede tomar. Los caminos posibles dependen de la cantidad máxima de pasos (o tiempo máximo)  $T_{max}$ . Este valor define cuántos caminos posibles existirán para la estimación. En primera instancia el investigador debe configurar y definir este  $T_{max}$ . Se debe tener en consideración que para 2 atributos y un  $T_{max}$  superior a 8 las matrices ya empiezan a ser muy grandes para que R pueda computarlas de forma eficiente.

Para conocer estas probabilidades de los caminos posibles se deben conocer todas las utilidades de todas las alternativas en cada posible camino de búsqueda. Luego, la estimación precisa del cálculo de utilidades y todas las probabilidades asociadas a las alternativas en cada paso  $t$ , dependiendo del camino que tomen todos los individuos dados los valores de los parámetros. El usuario codificador debe definir un  $T_{max}$ , el cual definirá la cantidad de caminos posibles y por consecuencia, el tamaño de las matrices (recordando que  $H = \sum_{\tau}^{T_{max}} K^{\tau}$ )

Por lo tanto, se define el cálculo de la verosimilitud implica tener en consideración las matrices indicadas en la Tabla 3.1.



Descripción matriz	Notación	Cantidad	Dimensión
Utilidad de alternativa	$V_{ni}$	$I$	$N \times H$
Utilidad posible futura de atributo $k$	$V_{sk}$	$K$	$N \times H$
Utilidad futura si individuo sigue observando	$V_{futura}$	1	$N \times H$
Utilidad presente si decido elegir	$V_{presente}$	1	$N \times H$
Probabilidad de seguir observando	$\mathbb{P}(s(t)   h)$	1	$N \times H$
Probabilidad de elegir	$\mathbb{P}(e(t)   h)$	1	$N \times H$
Probabilidad de elegir alternativa	$\mathbb{P}(i   h)$	$I$	$N \times H$
Probabilidad de observar atributo	$\mathbb{P}(k(t)   h)$	$K$	$N \times H$
Probabilidad de cada camino	$\mathbb{P}(h)$	1	$N \times H$
Matriz de herencia	$MH$	1	$H \times H$

Tabla 3.1: Matrices para la estimación del modelo. Probabilidades y utilidades por caminos de observación

Todo lo anterior, incluyendo el calculo de los parámetros  $\beta_k$ ,  $U_i^0$  (todas las alternativas menos 1),  $\alpha$ ,  $\delta$  y dos factores de escala de los tres posibles ( $\mu$ ,  $\mu_e$ ,  $\mu_s$ ) para la maximización de la función.

Para esto, se utiliza diferentes optimizadores del software R, que se explicarán en el Capítulo 4.

### 3.6. Diferencias teóricas con Decision Field Theory

Diferencias teóricas	Prototipo	Decision Field Theory
1. Distribución errores	• Valor Extremo	• Normal
2. Probabilidad observar atributo	• Dependiente del valor de la utilidad futura del atributo	• Depende del valor del parámetro del atributo
3. Regla decisión de seguir búsqueda	• Umbral interno dependiente de la cantidad de pasos	• Umbral interno y externo
4. Parámetros de proceso	• Convergencia $\alpha$	• Memoria $\phi_2$ y sensibilidad $\phi_1$
5. Interpretabilidad económica	• Se puede obtener relación entre parámetros, mas no es claro si es realmente el valor del tiempo	• No es posible
6. Función de decisión	• Utilidad	• Preferencia
7. Identificación de parámetros	• Sigue las reglas de modelos convencionales	• No existe claridad

Tabla 3.2: Diferencias teóricas del modelo prototipo y Decision Field Theory

### 3.7. Resumen ecuaciones del modelo prototipo

Se entrega un resumen de las ecuaciones del modelo, en caso de que en los siguientes capítulos se necesite hacer referencia a alguna de éstas.

1. Utilidad

$$U_{ni}^t = \alpha \cdot U_{ni}^{t-1} + (1 - \alpha) \cdot \beta_{k(t)} \cdot X_{nik(t)} + \varepsilon_{nik(t)} \quad (3.24)$$

2. Expresión de la probabilidad de alternativa y camino

$$\mathbb{P}(k(1), \dots, k(t), i) \quad (3.25)$$

3. Expresión de la probabilidad de atributo dado camino

$$\mathbb{P}(k(t) | h_i^{t-1}) = P(k(t) | s(t-1)) \cdot P(s(t-1) | h_i^{t-1}) \quad (3.26)$$

4. Expresión de la probabilidad de alternativa dado camino

$$\mathbb{P}(i | h_i^t) = P(i | e(t)) \cdot P(e(t) | h_i^{t-1}) \quad (3.27)$$

5. Utilidad presente

$$V_{presente}^t = \frac{1}{\mu_e} \cdot \ln \left( \sum_i^I e^{\mu_e \cdot V_i^t} \right) \quad (3.28)$$

6. Utilidad futura

$$V_{futura}^t = \frac{1}{\mu_s} \cdot \ln \left( \sum_k^K e^{\mu_s \cdot V_{sk}^{t+1}} \right) \quad (3.29)$$

7. Utilidad futura de un atributo

$$V_{sk}^{t+1} = \frac{1}{\mu_e} \cdot \ln \left( \sum_i^I e^{\mu_e \cdot V_{i|k}^{t+1}} \right) \quad (3.30)$$

8. Expresión probabilidad de seguir observando

$$\mathbb{P}(s(t) | h_i^t) = P(|V_{presente}^t - V_{futura}^t| \leq \delta_t) \quad (3.31)$$

9. Probabilidad de seguir observando

$$\mathbb{P}(s(t) | h_t^t) = \frac{e^{\mu \cdot V_{\text{presente}}^t}}{e^{\mu \cdot V_{\text{presente}}^t} + e^{\mu \cdot (V_{\text{futura}}^t + \delta_t)}} + \frac{e^{\mu \cdot V_{\text{futura}}^t}}{e^{\mu \cdot V_{\text{futura}}^t} + e^{\mu \cdot (V_{\text{presente}}^t + \delta_t)}} \quad (3.32)$$

10. Probabilidad de decidir elegir

$$\mathbb{P}(e(t) | h_t^t) = \frac{e^{\mu \cdot (V_{\text{futura}}^t + \delta_t)}}{e^{\mu \cdot V_{\text{presente}}^t} + e^{\mu \cdot (V_{\text{futura}}^t + \delta_t)}} + \frac{e^{\mu \cdot (V_{\text{presente}}^t + \delta_t)}}{e^{\mu \cdot V_{\text{futura}}^t} + e^{\mu \cdot (V_{\text{presente}}^t + \delta_t)}} \quad (3.33)$$

11. Probabilidad de observar atributo  $k$

$$\mathbb{P}(k(t+1) | s(t)) = \frac{e^{\mu_s \cdot V_{sk(t+1)}^{t+1}}}{\sum_{k'}^K e^{\mu_s \cdot V_{sk'}^{t+1}}} \quad (3.34)$$

12. Probabilidad de elegir alternativa  $i$

$$\mathbb{P}(i(t) | e(t)) = \frac{e^{\mu_e \cdot V_i^t}}{\sum_j^I e^{\mu_e \cdot V_j^t}} \quad (3.35)$$

13. Probabilidad de observar atributo  $k$  dado camino

$$\mathbb{P}(k(t) | h_t^{t-1}) = \underbrace{\frac{e^{\mu_s \cdot V_{sk(t)}^t}}{\sum_{k'}^K e^{\mu_s \cdot V_{sk'}^t}}}_{\mathbb{P}(k(t) | s(t-1))} \cdot \underbrace{\left( \frac{e^{\mu \cdot V_{\text{presente}}^{t-1}}}{e^{\mu \cdot V_{\text{presente}}^{t-1}} + e^{\mu \cdot (V_{\text{futura}}^{t-1} + \delta_t)}} + \frac{e^{\mu \cdot V_{\text{futura}}^{t-1}}}{e^{\mu \cdot V_{\text{futura}}^{t-1}} + e^{\mu \cdot (V_{\text{presente}}^{t-1} + \delta_t)}} \right)}_{\mathbb{P}(s(t-1) | h_t^{t-1})} \quad (3.36)$$

14. Probabilidad de elegir alternativa  $i$  dado camino

$$\mathbb{P}(i | h_t^t) = \underbrace{\frac{e^{\mu_e \cdot V_i^t}}{\sum_j^I e^{\mu_e \cdot V_j^t}}}_{\mathbb{P}(i | e(t))} \cdot \underbrace{\left( \frac{e^{\mu \cdot (V_{\text{futura}}^t + \delta_t)}}{e^{\mu \cdot V_{\text{presente}}^t} + e^{\mu \cdot (V_{\text{futura}}^t + \delta_t)}} + \frac{e^{\mu \cdot (V_{\text{presente}}^t + \delta_t)}}{e^{\mu \cdot V_{\text{futura}}^t} + e^{\mu \cdot (V_{\text{presente}}^t + \delta_t)}} \right)}_{\mathbb{P}(e(t) | h_t^t)} \quad (3.37)$$

15. Probabilidad de alternativa  $i$  para todo el proceso

$$\mathbb{P}(i) = \sum_{h_i^t}^{H^T} \mathbb{P}(i | h_i^t) \cdot \mathbb{P}(h_i^t) \quad (3.38)$$

16. Estimación: Log-verosimilitud

$$LL = \sum_n^N \sum_i^I y_{ni} \cdot \ln(\mathbb{P}(i)) \quad (3.39)$$

# Capítulo 4

## Ilustración, análisis y validación del modelo propuesto mediante casos simulados y reales

### 4.1. Evaluación y validez del modelo

En esta sección, se evalúa el comportamiento del modelo propuesto y cómo interaccionan sus parámetros. Se analiza, también, dónde pueden existir posibles falencias y dónde se comporta como se espera. Se estudia la función de utilidad, los parámetros propuestos y algunas probabilidades, para mejor entendimiento del prototipo propuesto.

#### 4.1.1. Función de utilidad

Como se definió en el capítulo anterior, la utilidad subyacente de este modelo cambia en el tiempo dependiendo de los atributos que observan en cada instante. Lo anterior busca replicar lo propuesto en el modelo Decision Field Theory, que la opinión (preferencia) de las alternativas que se observan es dinámica y dependen del momento en que se están observando. Por ejemplo, si tenemos una situación en que se elige en 20 segundos, puede que un individuo a los 5 segundos tenga una preferencia de alternativa diferente a la de los 20 segundos. Por lo tanto, se espera que este comportamiento se pueda explicar mediante la propuesta de modelo.

En la Figura 4.1 se muestra la evolución de utilidad de un individuo para ciertos parámetros. Se observa cómo la utilidad cambia a medida que pasa el tiempo y se observan los atributos secuencialmente, por lo tanto, la probabilidad de cada alternativa es, en efecto,

diferente en cada paso de observación  $t$ .

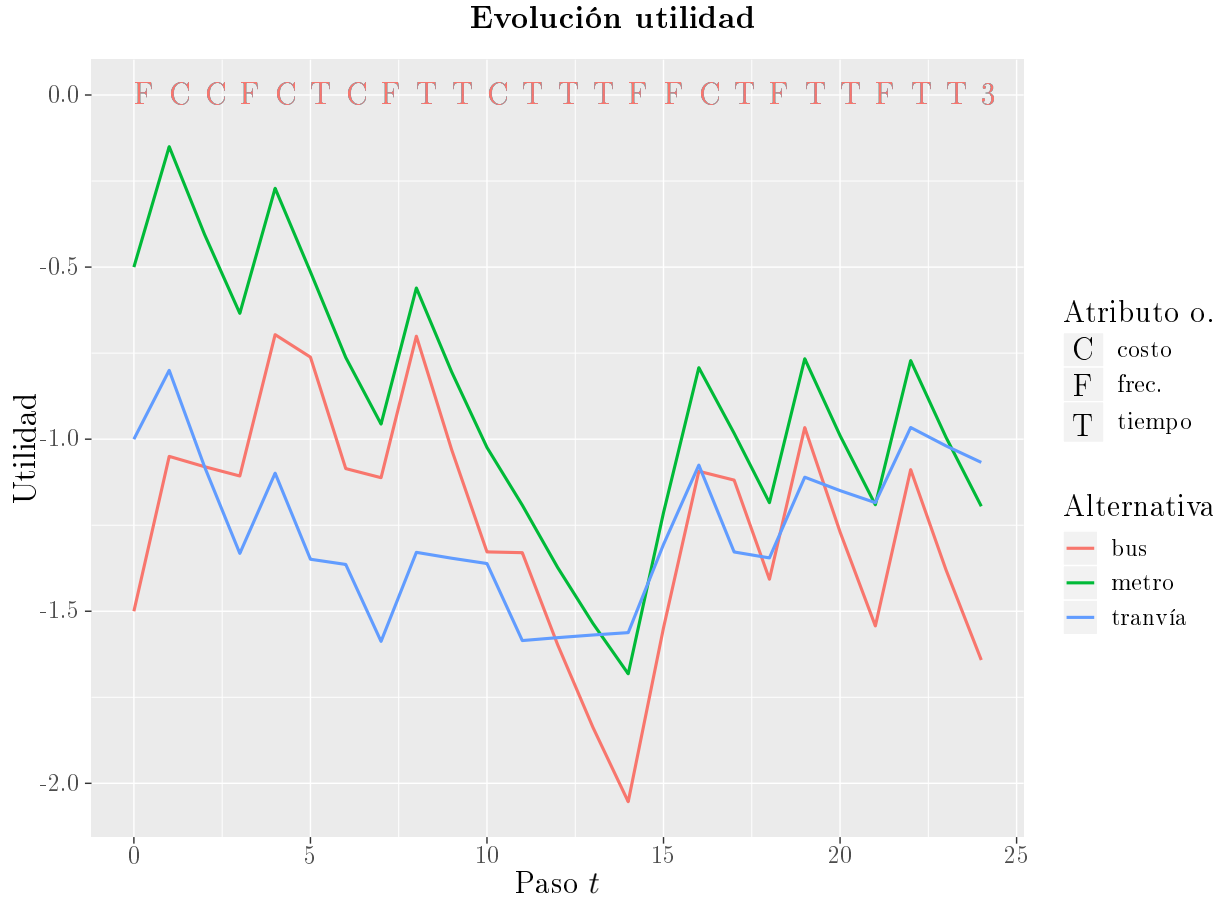


Figura 4.1: Evolución utilidad de individuo para 3 parámetros con 3 atributos

Finalmente, se comprueba que el modelo se comporta como se espera. No obstante, debido a que esta memoria es un prototipo de un modelo dinámico, en principio se espera que se comporte de manera menos precisa que Decision Field Theory.

#### 4.1.1.1. Parámetro $\alpha$ : peso del pasado

Es importante destacar en esta sección, que, luego de la formulación, y al empezar con el proceso de estudio de la estimación, se sospechó que el parámetro  $\alpha$  generaba que la función de utilidad y, por consecuencia, la función de verosimilitud, se volvieran no convexas. Esta limitación podría dificultar la estimación del modelo debido a que los algoritmos de optimización no podrían garantizar el que se alcanzó un óptimo global. Este problema ocurre con varios modelos como el DFT de Hancock, e incluso con el Nested Logit, y se manifiesta en que la solución encontrada dependería del punto de partida.

En efecto, se puede observar en la Ecuación 4.2 que al derivar las expresiones de utilidad para cada  $t$ , la segunda derivada de la utilidad con respecto a  $\alpha$  depende de  $\alpha$ , o algunos

de los  $\beta$ s (dependiendo de la forma de la utilidad, es decir, dependiendo de la cantidad de pasos  $T$  y del camino de búsqueda). No se entrará en detalle de cada una de las derivadas de las funciones utilidades, porque debido al espacio de posibles caminos, éstas pueden ser infinitas. Para ilustrar el problema, con la utilidad de  $t = 2$  ya es suficiente, ya que si ocurre para  $t = 2$  ocurrirá para  $t > 2$ , debido a la forma de la función de utilidad, ya que estas son dependientes de las utilidades anteriores.

Para esto, se ve el caso en que un individuo observa en  $t = 1$  el atributo  $k_1$ , y luego en  $t = 2$  observa otro atributo  $k_2$ , con esto, su utilidad será de la forma:

$$\begin{aligned}
V_{ni}^2 &= \alpha \cdot V_{ni}^1 + (1 - \alpha) \cdot \beta_{k_2} \cdot X_{nik_2} \\
V_{ni}^2 &= \alpha \cdot (\alpha \cdot V_{ni}^0 + (1 - \alpha) \cdot \beta_{k_1} \cdot X_{nik_1}) + (1 - \alpha) \cdot \beta_{k_2} \cdot X_{nik_2} \\
&\vdots \\
V_{ni}^2 &= \alpha^2 \cdot V_{ni}^0 + (1 - \alpha) \cdot \alpha \cdot \beta_{k_1} \cdot X_{nik_1} + (1 - \alpha) \cdot \beta_{k_2} \cdot X_{nik_2}
\end{aligned} \tag{4.1}$$

Y su segunda derivada:

$$\frac{\partial^2 V_{ni}^2}{\partial \alpha^2} = 2 \cdot V_{ni}^0 - 2 \cdot \beta_{k_1} \cdot X_{nik_1} \tag{4.2}$$

Como se observa, la segunda derivada de una de las utilidades de  $t = 2$  tiene un punto de inflexión que depende del valor de  $V_{ni}^0$ ,  $\beta_{k_1}$  y de  $X_{nik_1}$ . Entonces, la función de utilidad no es convexa en todo su conjunto, luego, por consecuencia, la función de verosimilitud tampoco lo es. Por lo tanto el parámetro  $\alpha$  causa no convexidad y el proceso de estimación del modelo se vuelve más complejo.

Por esto y por el costo computacional que requiere este modelo, se analizará el comportamiento de dos formas de estimación. La primera es asumiendo que se conoce el parámetro  $\alpha$ , y por lo tanto, el problema de no convexidad se elimina. Lo anterior supone que la estimación debe realizarse testeando valores del parámetro en cuestión. La segunda forma es estimando efectivamente este parámetro, utilizando el método convencional (método BFGS), pero también se probará la estimación mediante heurísticas de optimización más complejas, para poder analizar qué tan factible es recuperar el parámetro  $\alpha$ . En resumen se estimará mediante:

1. Estimación con parámetro  $\alpha$  conocido fijo y método BFGS y L-BFGS-B (menor costo computacional y problema convexo).
2. Estimación con parámetro  $\alpha$  desconocido y optimización mediante método BFGS y heurísticas de optimización (problema no convexo mejor abarcado, pero con mayor costo computacional asociado).

Para el segundo tipo de estimación se proponen dos heurísticas de optimización que según Mullen (2014) son opciones razonables para realizar la optimización de funciones no convexas en  $\mathbb{R}$ , en términos de reproducir resultados razonables y mejores tiempos de procesamiento. El primero es el paquete DEoptim (Ardia et al., 2011), el cual implementa un algoritmo evolucionario de derivadas (differential evolution algorithm) propuesto por Price et al. (2006), una estrategia similar a los algoritmos genéticos pero diseñado para optimización continua, lo que es útil para el problema de esta memoria. Por otro lado, el paquete Rmallschains (Bergmeir et al., 2016), utiliza un algoritmo para optimización continua usando "local search chains" (MA-LS-Chains por Molina et al., 2010).

#### 4.1.2. Parámetro $\delta_t$ : tolerancia

$\delta_t$  es el parámetro de tolerancia, que se define como la paciencia que el individuo tiene para seguir observando atributos y actualizando su utilidad. Éste es un parámetro dinámico, ya que el valor de  $\delta_t$  aumenta a medida que pasa el tiempo ( $\delta_t = \delta \cdot t^2$ ). Esto debido a que se considera que mientras los individuos observan, más impacientes se vuelven y más rápido quieren elegir. Esta lógica deriva de lo argumentado por Busemeyer y Townsend (1993), que señala que los individuos se auto-imponen implícitamente un tiempo límite para escoger. Además, esta lógica también puede ser explicada por el concepto de satisficing behaviour (Simon, 1957), en el cual los individuos al alcanzar un umbral interno proceden a escoger.

Es claro que cada individuo puede tener un valor diferente de este factor de tolerancia ( $\delta_{nt}$ ), sin embargo, para términos de esta memoria se generaliza para toda la población ( $\delta_{nt} = \delta_t$ ).

En términos poblacionales, la intuición señala que si el valor constante  $\delta$  aumenta, es decir, la tolerancia para cada paso  $t$  es menor, la desigualdad de la Ecuación 3.31 es cada vez más restrictiva, en otras palabras, los individuos prefieren no seguir observando, porque los cambios en su utilidad no son lo suficientemente grandes, entonces eligen antes. Luego, con todos los otros parámetros fijos se procede a calcular y comprobar esta hipótesis.

En la Figura 4.2 se puede observar el cambio en el promedio de pasos que les toma a los individuos elegir una alternativa dependiente del cambio del valor de  $\delta$  y, por consecuencia, de  $\delta_t$ . Se aprecia que a medida que  $\delta$  aumenta, el promedio de la cantidad de pasos de la población disminuye. Entonces, el modelo sigue la intuición respecto a este parámetro, ya que mientras más grande  $\delta$  es menos tiempo le toma a los individuos elegir, porque se vuelven más impacientes. Esto asumiendo mismo parámetro  $\delta$  para los individuos.



### Promedio pasos población v/s par. $\delta$

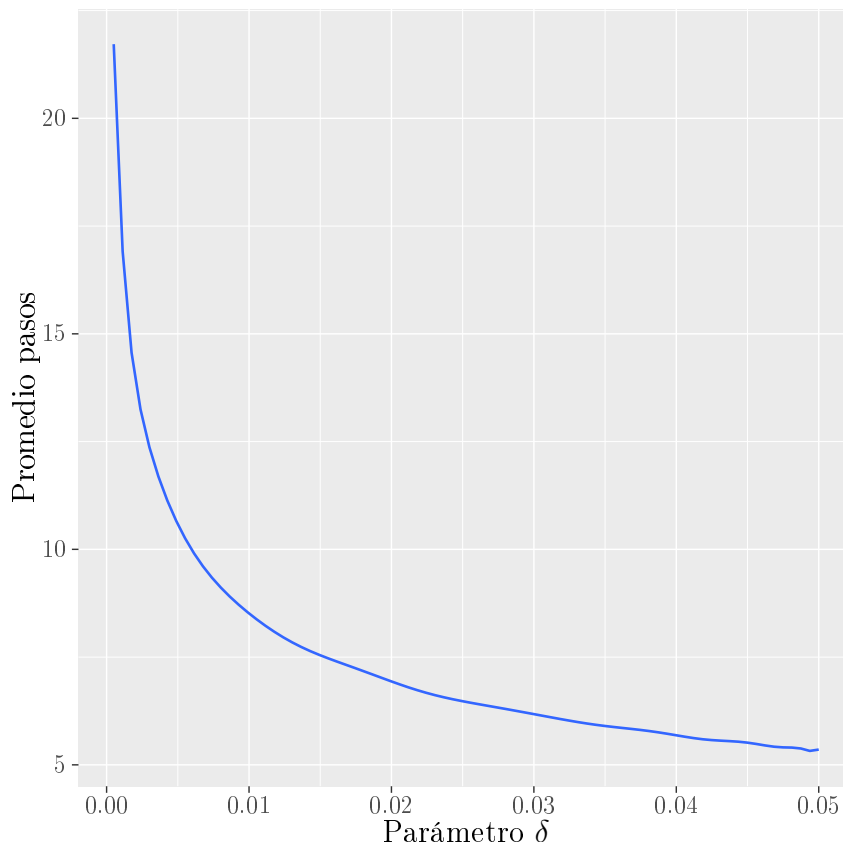


Figura 4.2: Efecto del valor del parámetro  $\delta$  en el promedio de pasos totales de la población

#### 4.1.3. Parámetro $U_i^0$ : Opinión inicial de alternativas

La opinión inicial es el parámetro que se considera la “constante” del modelo. Ésta, en teoría, representa la opinión que tiene el individuo pre-análisis de elección. Además, se intuye que también contiene factores no considerados en el modelo, como por ejemplo, la diferencia intrínseca entre las alternativas (como la constante específica de alternativas de MNL).

En esta sección, se probará la nominalidad del parámetro  $U_i^0$ , es decir, analizar el comportamiento si se trasladan todas las opiniones iniciales en un mismo valor cualquiera. Este análisis se realiza debido a que la estimación requiere fijar un parámetro en 0, para lograr la identificación del modelo. Entonces, si es que siempre debe estar un parámetro en 0, se debe estudiar si esta restricción cambia la descripción del comportamiento de los individuos. Esto es relevante debido a que en caso de que las utilidades iniciales no estén realmente alrededor del valor 0, puede que se no se esté no solo considerando la diferencia entre alternativas, sino la posición de los puntos, y por lo tanto, se estaría llevando a un sesgo de estimación.

En la Figura 4.3 se puede apreciar la evolución de utilidad de un individuo que debe

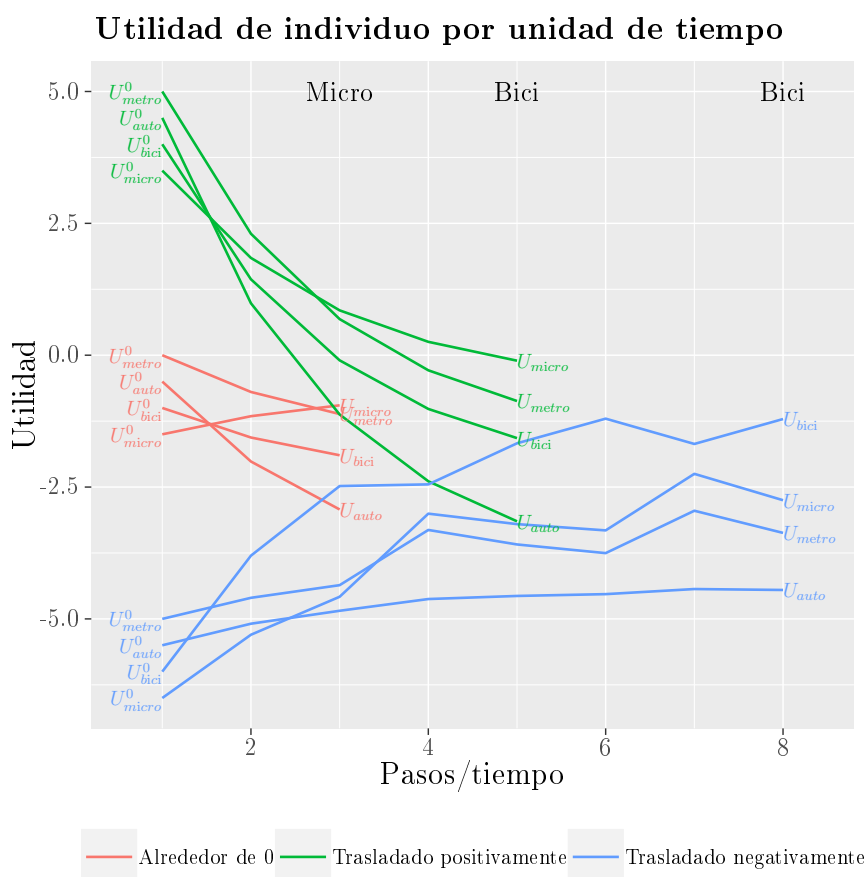


Figura 4.3: Evolución de utilidades para mismo individuo con 3 casos diferentes (puntos iniciales trasladados)

elegir entre 4 alternativas (bicicleta, automóvil, metro-tren, micro-bus) fijando el resto de los parámetros y fijando la semilla de aleatoriedad. Se puede apreciar que el comportamiento de las utilidades de todas las alternativas es distinta para los diferentes puntos iniciales y que el individuo elige diferentes alternativas dependiendo del punto de partida. Por lo tanto, se comprueba que al cambiar los puntos iniciales se tienen comportamientos diferentes, dada esta formulación del modelo.

Este problema se causa debido a que la expresión de probabilidad de seguir observando o elegir compara la utilidad futura con la utilidad presente. Pero al estar la utilidad presente multiplicada por  $\alpha$  en la expresión de la utilidad futura, ya no existe un factor común que mantenga nominalidad, y, por lo tanto, las probabilidades dependen de dónde se ubica el punto inicial de la utilidad de cada alternativa, no solo de su diferencia. Esto significa que, para que la estimación del modelo propuesto entregue una solución única, el investigador debiera conocer la utilidades iniciales, lo cual representa una clara limitación práctica del modelo prototipo.

Para dejar en claridad lo anterior, supongamos que el punto inicial para la estimación es

$A$  (caso 1) y el punto original es  $A + B$  (caso 2). Luego, por observar un atributo, la utilidad de  $t = 1$  se supone  $V_1^1 = A \cdot \alpha + C$  y la utilidad de observar otro atributo en  $t = 2$  es  $V_1^2 = V_1^1 \cdot \alpha + D$ . Entonces la probabilidad simplificada de elegir en el caso 1 es:

$$\mathbb{P}_1(e) = \frac{e^{A \cdot \alpha + C}}{e^{A \cdot \alpha + C} + e^{A \cdot \alpha^2 + C \cdot \alpha + D}} \quad (4.3)$$

Y de forma similar, la probabilidad para el caso trasladado es:

$$\mathbb{P}_2(e) = \frac{e^{(A+B) \cdot \alpha + C}}{e^{(A+B) \cdot \alpha + C} + e^{(A+B) \cdot \alpha^2 + C \cdot \alpha + D}} \quad (4.4)$$

Estas probabilidades no son iguales, pero se esperaría que lo fueran, ya que no debería importar la ubicación del punto inicial, sino que sólo la diferencia entre los puntos iniciales de las alternativas, para mantener la nominalidad. Luego, para que estas utilidades sean iguales, se debe multiplicar por  $\frac{1}{\alpha}$  la utilidad futura, así se elimina esta problemática.

En efecto, en la Ecuación 4.5 se observa que al multiplicar la utilidad futura por el factor  $\frac{1}{\alpha}$ , las probabilidades de elegir o no para ambos casos son iguales

$$\mathbb{P}_1(e) = \mathbb{P}_2(e) = \frac{e^0}{e^0 + e^{\frac{D}{\alpha}}} \quad (4.5)$$

En otras palabras, para que las utilidades presentes y futuras sean comparables y mantengan nominalidad, se debe traer a valor presente ( $\frac{1}{\alpha}$ ) la utilidad futura.

Finalmente, al modificar la formulación de la Ecuación 3.29 a la forma de la Ecuación 4.6 se arregla el problema de nominalidad, y, por lo tanto, ya no existe influencia al cambiar los puntos iniciales de lugar.

$$V_{futura}^t = \frac{1}{\mu_s} \cdot \ln \left( \sum_k^K e^{\mu_s \cdot V_{sk}^{t+1}} \right) \quad (4.6)$$

$$V_{futura'}^t = V_{futura}^t \cdot \frac{1}{\alpha}$$

Al generar datos los datos con esta modificación, efectivamente se tiene que el comportamiento de los individuos ya no es influenciado por dónde estén ubicados los puntos iniciales, sino que sólo influye la diferencia entre las utilidades, como se ve en la Figura 4.4.

Para el prototipo del modelo, esto no fue agregado a la formulación, ya que no es clara

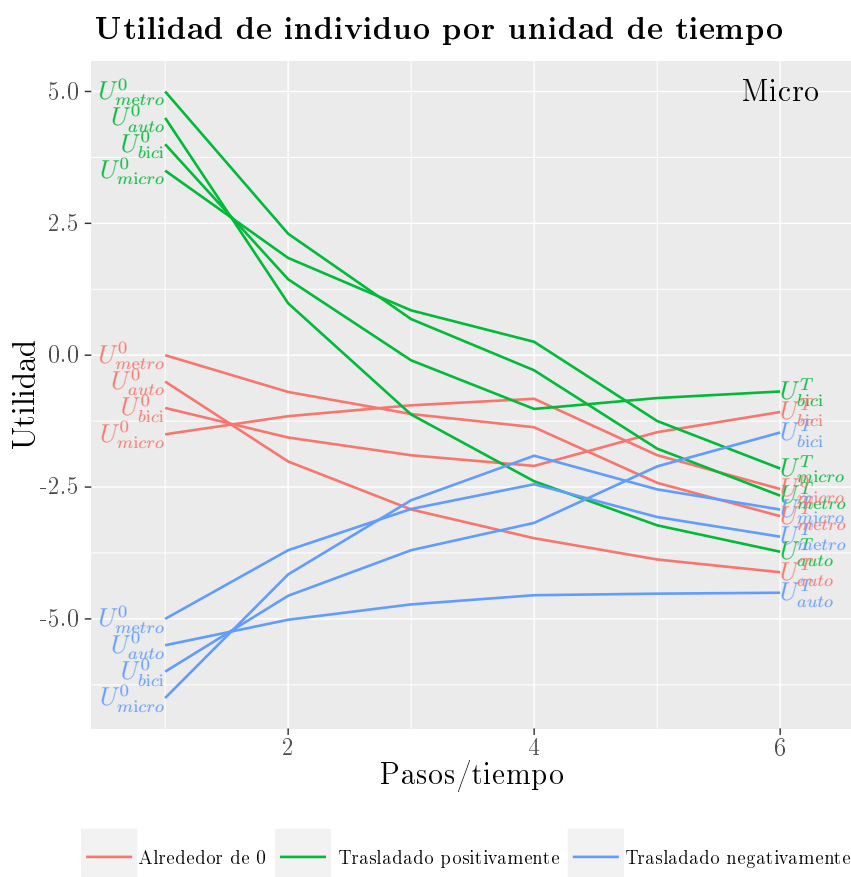


Figura 4.4: Evolución de utilidades para mismo individuo con 3 casos diferentes (puntos iniciales trasladados) con función utilidad futura modificada

la interpretación final y tampoco las implicancias de agregar el factor. Por lo tanto, parte de las mejoras futuras debe ser estudiar si incluir o no el factor  $\frac{1}{\alpha}$  en el cálculo de la utilidad futura para mantener nominalidad y no tener problemas con la estimación. Y en caso, de no incluirlo, buscar otro método para mejorar esta problemática.

#### 4.1.4. Probabilidad de observar atributos

Debido a la forma de la probabilidad de observar atributos  $\mathbb{P}(k)$  (3.34) se observa que, al existir mucha diferencia entre las utilidades entregadas por los atributos  $k$  en el siguiente paso  $t+1$ , las probabilidades son poco realistas. Por ejemplo, cuando existe un atributo que tiene un efecto positivo y otro que tiene un efecto negativo en la utilidad, la probabilidad de observar el atributo positivo siempre será mayor a mirar la del atributo negativo, ya que la utilidad siempre crecerá con el positivo, mientras que siempre disminuirá con el negativo. Luego, el individuo siempre tendrá una probabilidad mayor de observar dicho atributo positivo, lo que es poco realista. Lo anterior es independiente de la importancia real que tiene el atributo, por lo tanto, el modelo, en esta probabilidad, tiene problemas de formulación.

## 4.2. Análisis del método de estimación: Monte Carlo

En esta sección se realizará un experimento de Monte Carlo, en donde se generarán los datos de elección utilizando el modelo. Después, se estimarán los parámetros con estos datos para poder verificar si el modelo es identificable (1), si se pueden recuperar los parámetros poblacionales (2) y para comparar su desempeño con otros modelos existentes (3).

Para la estimación, como se declaró en 3.4.2, se recurrirá a un análisis extenso de cómo estimar el modelo. Esto, ya que, como se mencionó en algunos puntos de la sección anterior, existen problemas con algunos parámetros ( $\alpha$  y  $U_i^0$ ) que deben ser estudiados. De la misma manera, se hará un análisis del costo computacional asociado a la cantidad de pasos máximos que se deben tener, ya que existe un límite de este valor para el cálculo de la estimación. Finalmente, se elegirá una opción de estimación para compararla con otros modelos.

Es importante señalar que para la estimación se supone que solamente se tienen los mismos datos disponibles que para cualquier otro modelo convencional, por lo tanto sólo se conocen los valores de los atributos y las elecciones. En caso de que se conozca otra información, esto se indicará debidamente, por ejemplo, en estimación con camino conocido o tiempo conocido, la que podría ser obtenida a través de encuestas PD en línea y/o de experimentos de eye-tracker.

### 4.2.1. Datos

La situación en que se encontrará este experimento de Monte Carlo es un contexto ficticio de una encuesta de preferencia declarada de 4 alternativas de movilización diaria dentro de la ciudad (bicicleta, metro-tren, micro-bus y automóvil). Se le presenta al encuestado solamente 2 atributos para cada alternativa: tiempo de viaje y costo del viaje (por simpleza del primer modelo a probar). El encuestado conoce los valores exactos de los atributos que se le presentan y debe elegir en base a estos atributos y en una ventana de tiempo definida previamente, una de las 4 opciones.

Los datos de los atributos se generan con una distribución uniforme. El precio del viaje en bicicleta fluctúa entre \$0 y \$100 y el precio de utilizar automóvil varía entre \$0 y \$1800. Para las otras dos opciones existen solo dos posibles valores, para micro-bus es \$230 o \$700 y para el de metro-tren es \$230 o \$800. Luego, el tiempo de viaje para bicicleta es de 10 a 20 minutos, el del automóvil de 2 a 25 minutos, el del metro-tren de 4 a 15 minutos, y finalmente el de micro-bus de 3 a 22 minutos. Se supone una situación simple en donde todos los individuos tienen acceso a todos los modos de transporte. Se generan los atributos para

1 situación de elección con 4000 individuos.

Simbolo	Parametro	Valor (general)
$\beta_t$	Tiempo de viaje	-2.0
$\beta_c$	Costo del viaje	-0.5
$U_{bici}^0$	Utilidad inicial bicicleta	-1.0
$U_{micro}^0$	Utilidad inicial micro-bus	-0.5
$U_{metro}^0$	Utilidad inicial metro-tren	0
$U_{auto}^0$	Utilidad inicial automóvil	-0.5
$\delta_t$	Tolerancia	0.1
$\alpha$	Factor temporal	0.6
$\mu$	Escala	0.1
$\mu_e$	Escala	0.1
$\mu_s$	Escala	0.1

Tabla 4.1: Valor parámetros para generación de datos

El valor de los parámetros utilizados para generar los datos se encuentran en la Tabla 4.1. Cualquier cambio en estos parámetros o algún supuesto que se tenga se mencionará debidamente.

#### 4.2.2. Estimación con camino conocido

La primera estimación a analizar es la estimación de camino conocido. Esta supone que se tiene como información lo observado por los individuos en cada paso  $t$ . Esto supone entonces que el parámetro  $\delta$  y los parámetros de escala ( $\mu_s$ ,  $\mu_e$  y  $\mu$ ) no son estimados, ya que al conocer el camino de búsqueda del individuo no se precisa de las probabilidades de decisión entre nodos. El resto de los parámetros se estima diferenciando entre la estimación con  $\alpha$  conocido fijo y con  $\alpha$  desconocido.

Dentro del análisis de esta sub-sección y de la siguiente, se realizarán 20 simulaciones y se estimarán los parámetros, esto con el fin de tener una noción de qué tanto cambian los resultados si se modifican las variables. Se mostrarán los resultados de estas simulaciones mediante gráficos boxplot, los cuales reportan, según parámetro (o métrica), la distribución de los valores entregados en cada simulación. Éstos gráficos muestran la distribución de las variables, donde las barras representan el valor del primer cuartil al tercer cuartil, la línea que corta la barra es la mediana (segundo cuartil), y las líneas que salen de la barra son los valores extremos, y los puntos son los outliers o valores “raros”. Esta forma de presentar los resultados es para analizar cómo se distribuyen los datos de las simulaciones de forma visual.

### 4.2.2.1. $\alpha$ conocido

Se realizan 20 simulaciones de elecciones generadas aleatoriamente con el modelo propuesto. Luego, cada simulación es estimada con el modelo, suponiendo que se conoce lo que cada individuo observó, y que el parámetro  $\alpha$  también es conocido.

**Boxplot media parámetros estimados (20 simulaciones)**

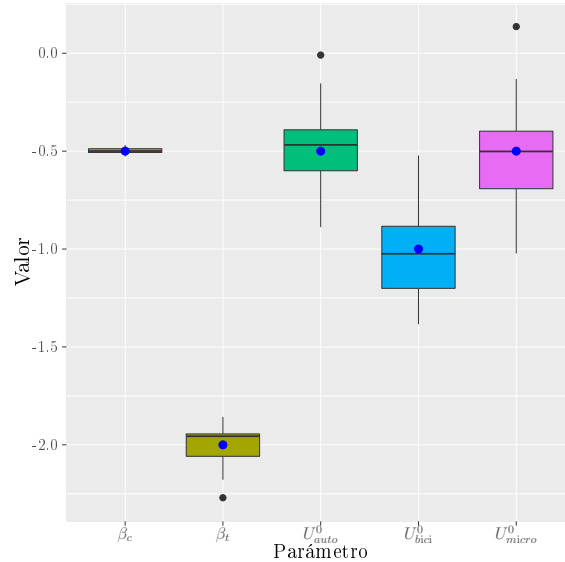


Figura 4.5: Valor media parámetros estimados de 20 simulaciones y valores reales (puntos azules) con  $\alpha$  fijo

En la Figura 4.5 se muestran los resultados de estimación del valor de cada parámetro para estas simulaciones. Se puede apreciar que todos los parámetros que se estiman aparentemente se encuentra dentro de una vecindad del punto real (punto azul). El parámetro  $\beta_c$  es el parámetro que se estima de manera más “consistente”, ya que todos los valores son muy similares y no reportan valores extremos ni outliers.

### Boxplot p-valor parametros (20)

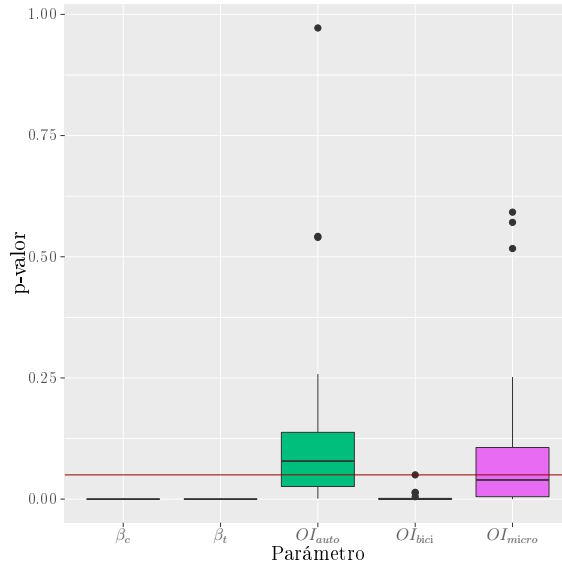


Figura 4.6: t-test significancia a parámetros estimados de 20 simulaciones y nivel de significancia 0.05 (línea roja)

Además, se aprecia que en estas 20 simulaciones, 3 de los 5 parámetros siempre son estadísticamente significativos con 95 % de confianza (4.6). Es discutible si los otros dos parámetros  $U_{auto}^0$  y  $U_{micro}^0$  se pueden analizar de esta manera ya que su valor real (0.5) es cercano a 0, y como se mencionó, existen problemas de estimación para las opiniones iniciales  $U_i^0$ .

### p-valor para test de media igual a valor real (20 simulaciones)

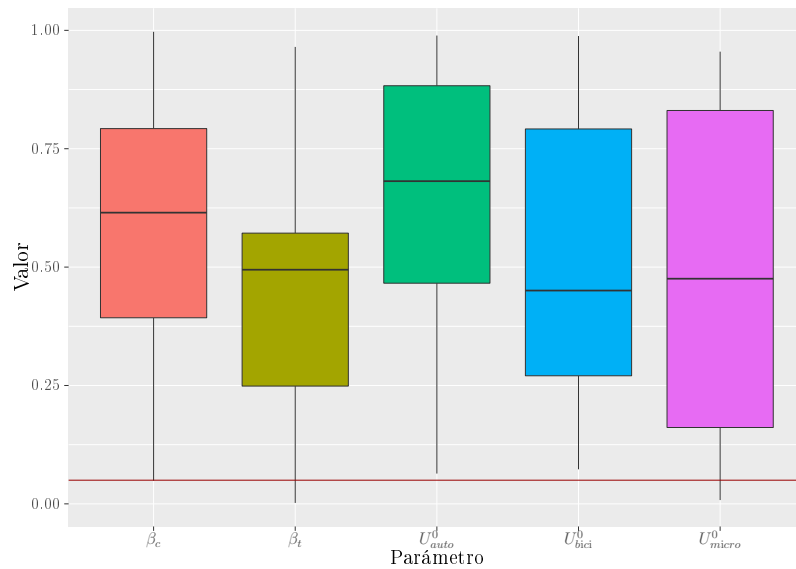


Figura 4.7: Boxplot p-valor para test de media igual a valor real de 20 simulaciones. Línea roja es p-valor = 0,05



Luego, al analizar si los valores estimados en las 20 simulaciones son estadísticamente iguales al valor real con 95% de confianza, se observa en la Figura 4.7 que para todos los parámetros la mediana del p-valor es considerablemente mayor a 0.05, por lo que no se puede rechazar la hipótesis nula de parámetros igual al valor real. En otras palabras, se puede interpretar que el modelo con camino conocido sí recupera el valor real de los parámetros

#### 4.2.2.2. $\alpha$ desconocido

Es importante destacar que para este caso, la estimación de  $\alpha$  no precisó del uso de heurísticas, ya que se probó estimar con 5 puntos de partida diferentes para una simulación.

	$PI_1$	$PI_2$	$PI_3$	$PI_4$	$PI_5$
$\beta_t$	0.1	0.1	0.1	-5.0	5.0
$\beta_c$	0.1	0.1	0.1	-5.0	5.0
$U_{bici}^0$	0.1	0.1	0.1	-5.0	5.0
$U_{micro}^0$	0.1	0.1	0.1	-5.0	5.0
$U_{auto}^0$	0.1	0.1	0.1	-5.0	5.0
$\alpha$	0.5	0.1	0.9	0.9	0.1

Tabla 4.2: Puntos iniciales diferentes para estimación con parámetro  $\alpha$

Los puntos utilizados se pueden observar en la Tabla 1, en donde los primeros 3 conjuntos tienen los valores iniciales estándar que se utilizan generalmente en el modelo, solamente variando el valor del parámetros  $\alpha$ , por otro lado, para los conjuntos 4 y 5, se variaron todos los puntos iniciales. El último conjunto es el conjunto del punto anterior, en donde se asume que se conoce el parámetro  $\alpha$ .

	Real	$PI_1$	$PI_2$	$PI_3$	$PI_4$	$PI_5$	$\alpha$ conocido
$\beta_t$	-2.000	-1.968	-1.968	-1.968	-1.968	-1.968	-1.954
$\beta_c$	-0.500	-0.475	-0.475	-0.475	-0.475	-0.475	-0.471
$U_{bici}^0$	-1.000	-1.099	-1.098	-1.098	-1.098	-1.099	-1.136
$U_{micro}^0$	-0.500	-0.991	-0.989	-0.991	-0.991	-0.990	-1.022
$U_{auto}^0$	-0.500	-0.463	-0.463	-0.463	-0.463	-0.463	-0.479
$\alpha$	0.600	0.607	0.607	0.607	0.607	0.607	-
LL		4560.643	4560.643	4560.643	4560.643	4560.643	4560.643

Tabla 4.3: Estimación camino conocido para diferentes puntos iniciales con parámetro  $\alpha$

En la Tabla 4.3 se aprecian los resultados de la estimación, en donde al tener un camino conocido se puede estimar con optimizador convexo. (BFGS).

### 4.2.3. Estimación con cantidad de pasos de elección conocida

El segundo test de estimación es el de estimar el modelo suponiendo que se conoce el la cantidad de pasos que le tomó a cada individuo realizar la elección. Al igual que en la subsección anterior, el parámetro  $\delta$  y los parámetros de escala tampoco pueden ser estimados, ya que no son necesarios para el calculo de la log-verosimilitud.

#### 4.2.3.1. $\alpha$ conocido

Para testear el modelo suponiendo que el parámetro  $\alpha$  es conocido, se realizan 20 simulaciones, y luego se estiman los parámetros.

**Boxplot media parámetros estimados 20 simulaciones (cantidad pasos conocido,  $\alpha$  conocido)**

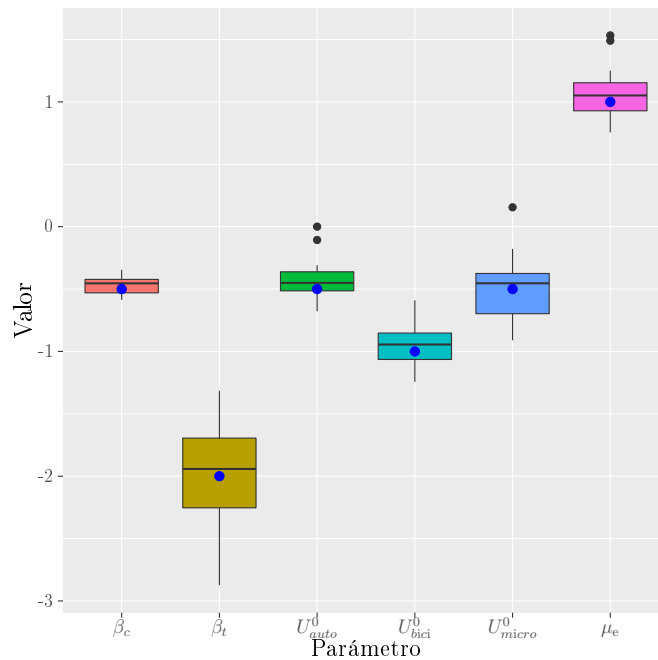


Figura 4.8: t-test significancia a parámetros estimados de 20 simulaciones y nivel de significancia 0.05 (línea roja) con  $\alpha$  desconocido

Se puede observar de la Figura 4.8 que los parámetros estimados se encuentran en un vecindario cercano al valor real, en donde el valor real se encuentra dentro del vecindario del cuartil 1 al cuartil 3. Sin embargo, al realizar un test de hipótesis de si los parámetros son iguales a los valores reales, se obtiene que no se puede rechazar la hipótesis nula al 95 % de confianza (Figura 4.9). Por lo tanto, se debe realizar un análisis de la relación entre parámetros, ya que debido a la forma del logit, los factores de escala probablemente son diferentes a 1, y por esto no se estaría recuperando el valor real de los parámetros.

**p-valor para test de medias de las 20 simulaciones (cantidad pasos conocido,  $\alpha$  conocido)**

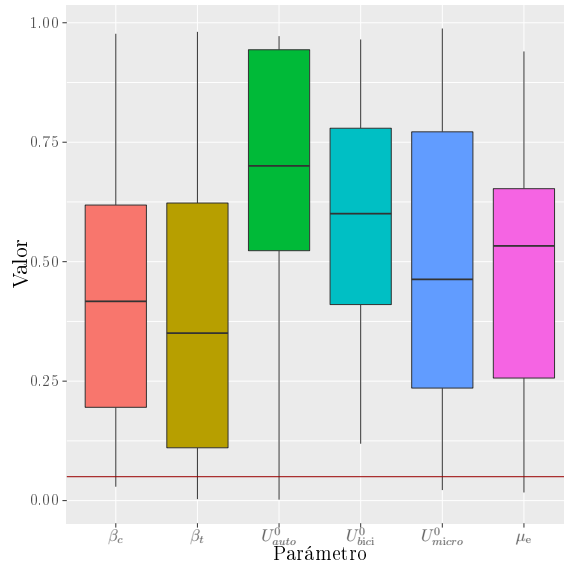


Figura 4.9: Boxplot p-valor para test de media igual a valor real de 20 simulaciones. Línea roja es p-valor = 0,05

Luego, en la Figura 4.9, se observa que la relación entre los parámetros en todos los casos no se rechaza la hipótesis nula del parámetro igual al valor real. Se exceptúan los casos de valores extremos de los parámetros  $\beta_t$  y  $U_{micro}^0$  que tienen un p-valor menor a 0.05. Por lo tanto, el modelo con tiempo conocido y parámetro  $\alpha$  conocido puede recuperar el valor de los parámetros en casi todas las simulaciones realizadas.

Finalmente, respecto a la significancia de los parámetros, se observa que éstos son estadísticamente significativos (tiempo y costo). Para los puntos iniciales ( $U^0$ ) no se puede realizar esta conclusión, ya que se necesita ajustar el modelo para mejorar la estimación de puntos iniciales.

### Boxplot nivel de significancia de parámetros en 20 simulaciones (cantidad pasos conocido, $\alpha$ conocido)

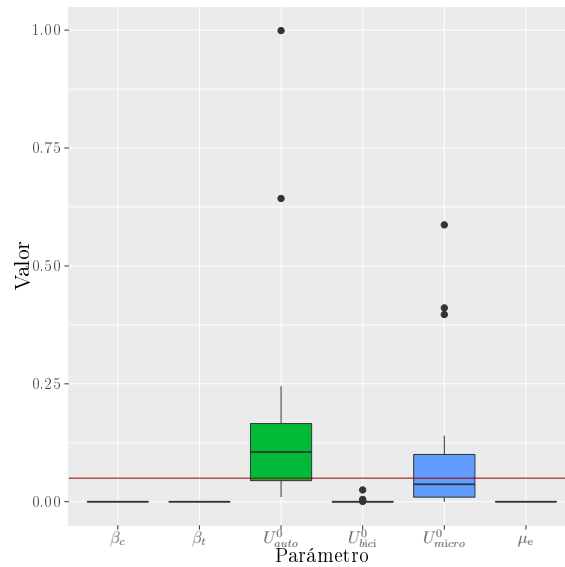


Figura 4.10: Valor media parámetros estimados de 20 simulaciones y valores reales (puntos azules)

#### 4.2.3.2. $\alpha$ desconocido

Para este análisis se realizó una simulación, a diferencia del análisis anterior, ya que la estimación con optimizador global DEoptim toma mucho más tiempo de estimación. Se realizaron dos estimaciones, primero con optimizador global DEoptim y luego con optimizador local BFGS utilizando los resultados de DEoptim como puntos iniciales, para poder obtener el valor de la desviación estándar, ya que el optimizador DEoptim no entrega hessiano.

Luego, en la Tabla 4.4 se muestran los resultados de la simulación. En donde se puede apreciar que el parámetro  $U_a^{0uto}$  es el único no estadísticamente significativo al 95% de confianza. Además, todos los parámetros a excepción de  $\mu_e$  y  $\beta_t$  son estadísticamente iguales al valor real (no se puede rechazar hipótesis nula al 95%. Por lo tanto, el modelo puede recuperar casi satisfactoriamente los parámetros, sin embargo debe indagarse y estudiarse en trabajos futuros el poder recuperar todos los parámetros correctamente.

	Estim	Std. Error	p-valor	p-valor valor real
$\beta_t$	-1.387	0.229	0.000	0.008
$\beta_c$	-0.416	0.079	0.000	0.290
$U_{bici}^0$	-1.736	0.503	0.001	0.143
$U_{micro}^0$	-1.243	0.432	0.004	0.086
$U_{auto}^0$	-0.710	0.368	0.054	0.569
$\mu_e$	1.164	0.167	0.000	0.000
$\alpha$	0.509	0.063	0.000	0.150

Tabla 4.4: Resultados de la estimación de los parámetros de una simulación para cantidad de pasos conocida y parámetro  $\alpha$  desconocido

#### 4.2.4. Estimación convencional: probabilidades latentes de caminos de búsqueda

Para el tercer análisis de estimación, se pretende realizar el cálculo de la log-verosimilitud suponiendo que solamente se conocen las elecciones y los valores de los atributos para los individuos. Esta forma se supone la forma tradicional de estimar el modelo, ya que la gran mayoría de las bases de datos existentes en la literatura contienen sólo esta información.

Dicho esto, se analizará la estimación dentro de todas sus aristas, ya que existen algunos supuestos que se deben clarificar antes de empezar a comparar el modelo con otros.

Los datos generados para todo el análisis de esta sub-sección son generados con sólo una semilla aleatoria, para poder realizar análisis comparables entre las formas de estimar y los supuestos que se realizan.

Bicicleta	Micro-bus	Metro-tren	Automóvil
1289	867	1020	824

Tabla 4.5: Cantidad de veces que las alternativas fueron escogidas

En la Tabla 4.5 y en la Figura 4.11 se observan la cantidad de veces que cada alternativa fue elegida en la población, y la la distribución de la cantidad de etapas que le tomó a los individuos realizar sus elecciones.

### Histograma pasos realizados por población

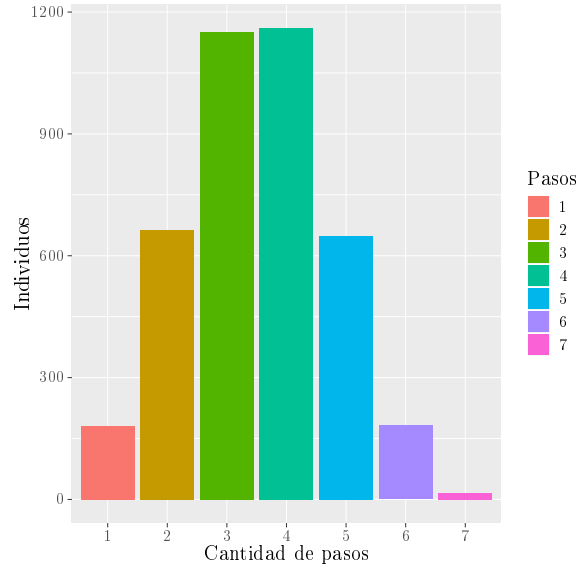


Figura 4.11: Cantidad de pasos realizados de la población

#### 4.2.4.1. Factores de escala $\mu$ , $\mu_e$ , $\mu_s$

Al considerar este modelo 3 factores de escala, 2 de estos deben ser estimados. El método de obtención de los valores de los parámetros se realiza mediante una estimación simultánea, lo que genera que el modelo se vuelva un modelo no convexo (Ben-Akiva et al., 1985), por lo tanto éste debe ser estimado secuencialmente para que esto no ocurra.

Esto no fue considerado en la formulación del método codificado de estimación, por lo tanto,

	PP $\mu$ general	PP1	PP2	PP3	P44
$\beta_t$	0.1	0.1	-2	0.1	0.1
$\beta_c$	0.1	0.1	-2	0.1	0.1
$U_{bici}^0$	0.1	0.1	-2	0.1	0.1
$U_{micro}^0$	0.1	0.1	-2	0.1	0.1
$U_{auto}^0$	0.1	0.1	-2	0.1	0.1
$\delta$	0.001	0.001	0.001	0.001	0.001
$\mu$	—	1	1	0.01	10
$\mu_e$	—	1	1	0.01	10

Tabla 4.6: Puntos iniciales para estimación con factores de escala

al tratar de estimar el modelo con los 2 parámetros de escala éstos generan ruido en los resultados, ya que al cambiar el punto inicial, cambia el valor final de los parámetros.

En efecto, en la Tabla 4.7 se observan los resultados de la estimación utilizando diferentes puntos iniciales (Tabla 4.6). Se aprecia que para diferentes puntos iniciales se tienen diferentes

Parámetros	$\mu$ general		PP1		PP2		PP3		PP4	
	media	t-test	media	t-test	media	t-test	media	t-test	media	t-test
$\beta_t$	-1.67	0.00	-1.42	—	-1.59	0.00	-1.45	0.00	-0.21	—
$\beta_c$	-0.53	0.00	-0.46	—	-0.53	0.00	-0.48	0.00	-0.05	—
$U_{bici}^0$	-1.41	0.14	-0.88	0.00	-2.77	0.00	-0.92	0.05	-0.12	—
$U_{micro}^0$	-0.59	0.22	-0.28	0.11	-1.13	0.06	-0.30	0.21	-0.05	0.00
$U_{auto}^0$	-0.88	0.08	-0.51	0.04	-2.46	0.00	-0.53	0.06	-0.04	0.01
$\delta$	0.04	0.36	0.10	—	0.01	0.47	0.55	—	0.09	0.00
$\mu$	N.E	—	1.00	—	1.30	0.77	0.17	—	10.00	—
$\mu_e$	N.E	—	1.31	—	0.95	0.00	1.27	0.00	9.90	—
tiempo	223.12		300.04		288.61		300.63		274.56	
$LL$	4959.97		4958.55		4961.66		4958.53		4972.07	

Tabla 4.7: Resultados estimación con factores de escala para diferentes puntos iniciales valores de los parámetros estimados. En la primera columna se tienen los resultados al estimar el modelo con todos los factores de escala fijados en 1.

Para simplificación, se asumirá de aquí en adelante que  $\mu = \mu_e = \mu_s$ , lo que simplifica el problema de codificación debido a que la estimación es una estimación simultanea. Futuros trabajos deben buscar la simplificación de este problema, agregando un método de estimación secuencial. Otras opciones pueden ser eliminando expresiones que consideren agregar factores de escala nuevos, o crear un nuevo método de estimación que pueda atrapar el efecto de los factores de escala sin caer en este problema de optimización no convexo.

#### 4.2.4.2. Costo computacional: tiempo máximo de elección y de estimación

Como se ha mencionado anteriormente, el hecho de estimar el modelo integrando respecto a todos los caminos posibles que el individuo puede realizar genera un problema matricial. Por lo tanto, para la estimación se debe tener en consideración este problema, ya que el proceso de optimización para recuperar los parámetros puede tomar muchos minutos o hasta horas si se asumen muchos pasos de observación. Incluso puede llegar a no poder calcularse, debido a las limitaciones del software R o del mismo computador.

Para este análisis es necesario diferenciar el tiempo máximo de elección y el de estimación. El tiempo máximo de elección es el tiempo total del individuo que más demora en analizar su elección. Es posible que este tiempo máximo de elección sea impuesto. Por otro lado, el tiempo máximo de estimación, es el que se utiliza para realizar los cálculos de las matrices de utilidades y probabilidades en la estimación. Como se mencionó en la Sección 3.5, se debe tener en consideración, que para la maximización de la log-verosimilitud, se necesita realizar una optimización en una función con una gran cantidad de matrices, que crecen de acuerdo al

valor del tiempo máximo ( $T_{max}$ ). Por ello, el proceso de optimización de la función se vuelve más complejo, ya que mientras más grande sea  $T_{max}$ , más se exigirá al ordenador que calcule. En cada  $t$  se agregan al modelo  $K^t$  opciones de caminos posibles a observar, por lo que se debe ser precavido en qué  $T_{max}$  se debe definir. Futuros trabajos deben buscar la forma de poder incorporar una estimación matricial más rápida y factible, ya que pueden existir casos que el tiempo promedio de observación sea mucho mayor al  $T_{max}$ , por lo que la estimación podría verse muy afectada, ya que no estaría representando el proceso real que ocurre.

Se estudiará el efecto que tiene el estimar con diferentes valores de  $T_{max}$ , para poder analizar qué tanto afecta el no utilizar la cantidad de pasos necesaria.

Para el siguiente análisis, se estudiará el cambio del valor de los parámetros y del valor de la log-verosimilitud en función del valor de  $T_{max}$  suponiendo que se conoce el valor de  $\alpha$ . Además, ¿qué sucede cuando el tiempo promedio de elección de los individuos supera el tiempo máximo  $T_{max}$  definido para la estimación? Para responder esta pregunta, se hará un análisis de qué sucede si  $T_{max}$  es menor, igual o mayor al tiempo máximo de elección entre los individuos, además de observar el costo computacional asociado a aumentar el  $T_{max}$ .

A continuación se presentan las elecciones generadas y el tiempo que le tomó a los individuos realizar estas elecciones, en donde el tiempo máximo de elección fueron 7 pasos y el promedio 4.53 pasos.

Luego, los resultados de la estimación con  $\alpha$  conocido se pueden observar en la Tabla 4.8. De la Tabla 4.8 se puede observar que a medida que se aumenta el espacio de caminos

$T = 7$ real	Real	$T_{max} = 1$	$T_{max} = 2$	$T_{max} = 3$	$T_{max} = 4$	$T_{max} = 5$	$T_{max} = 6$	$T_{max} = 7$	$T_{max} = 8$	$T_{max} = 9$
$\beta_t/\beta_c$	4.00	5.329	4.200	3.455	3.571	3.668	3.679	3.678	3.680	3.678
$\beta_t$	-2.00	-4.325	-2.582	-2.001	-2.039	-2.099	-2.107	-2.107	-2.108	-2.107
$\beta_c$	-0.50	-0.812	-0.615	-0.579	-0.571	-0.572	-0.573	-0.573	-0.573	-0.573
$U_{bici}^0$	-1.00	-0.401	-0.438	-0.681	-0.706	-0.668	-0.661	-0.662	-0.661	-0.662
$U_{micro}^0$	-0.50	-0.117	-0.133	-0.197	-0.203	-0.194	-0.191	-0.192	-0.192	-0.192
$U_{auto}^0$	-0.50	-0.140	-0.011	0.050	0.031	0.020	0.019	0.018	0.018	0.018
$\delta$	0.09	0.001	0.001	0.038	0.115	0.138	0.141	0.140	0.141	0.140
LL	0.00	4973.616	4961.620	4960.499	4960.660	4960.722	4960.726	4960.726	4960.726	4960.726
TE	0.00	3s	7s	17s	46s	1m 49s	3m 49s	8m 29s	12m 49s	34m 35s

Tabla 4.8: Estimación parámetros para tiempos máximos de estimación diferentes. Siendo  $T=7$  el tiempo máximo de elección de los individuos.

posibles, aumenta el tiempo de estimación, lo cual es esperable. Además, es interesante que para  $T_{max} = 3$  y  $T_{max} = 4$  la log-verosimilitud entrega un mejor ajuste siendo que  $T_{max} = 7$  es el valor que debería utilizarse (en teoría). Además, se observa que para  $T_{max} = 5, 6, 7, 8$  y  $9$  los parámetros estimados son similares. Lo anterior puede deberse a que desde  $T_{max} = 5$  se supera el promedio de la población, por lo que al estimar con este valor puede que ya se esté



representado a gran parte de la población.

$T = 7$ real	$T_{max} = 1$	$T_{max} = 2$	$T_{max} = 3$	$T_{max} = 4$	$T_{max} = 5$	$T_{max} = 6$	$T_{max} = 7$	$T_{max} = 8$	$T_{max} = 9$
$\beta_t$	0.212	0.303	0.370	0.389	0.359	0.356	0.358	0.357	0.358
$\beta_c$	0.048	0.054	0.069	0.072	0.068	0.068	0.068	0.068	0.068
$U_{bici}^0$	0.086	0.138	0.328	0.245	0.216	0.214	0.215	0.215	0.215
$U_{micro}^0$	0.079	0.150	0.269	0.205	0.190	0.188	0.189	0.189	0.189
$U_{auto}^0$	0.058	0.140	0.221	0.170	0.166	0.165	0.165	0.165	0.165
$\delta$	NA	0.386	0.197	0.157	0.139	0.137	0.138	0.138	0.138

Tabla 4.9: Desviaciones estándar estimación para Tiempos máximos diferentes

#### 4.2.4.3. Parámetro $\alpha$

Debido a que el parámetro  $\alpha$  convierte a la función en una función no convexa en todo su dominio, se propone realizar su estimación probando el  $\alpha$  como parámetro fijo, analizando qué tanto cambia si se propone fijar el parámetro y estimarlo con valores diferentes en comparación a estimarlo con heurísticas. Se realiza este análisis para contrastar los resultados de los diferentes modelos estimados con  $\alpha$  diferentes fijos. Se busca ver cual es el impacto de fijar el parámetro a uno que no es el real.

La Tabla 4.10 muestra los valores de estimación del modelo para diferentes valores de  $\alpha$ . De estas estimaciones, se aprecia que para los valores más bajos de  $\alpha$  (0.05 y 0.10) el modelo entrega valores de los parámetros muy alejados del valor real. Además, para  $\alpha = 0,80$  se tiene el valor más bajo de log-verosimilitud, lo que nos indica que el modelo no está ajustando mejor de acuerdo al valor real del parámetro. La relación entre los parámetros de los atributos ( $\beta_t/\beta_c$ ) que más se acerca al valor real (4) es para  $\alpha = 0,50$ , por lo tanto, en este ámbito tampoco se está calculando correctamente el modelo. Por último, se observa que para todos los valores de  $\alpha$ , la relación entre los parámetros se encuentra en una vecindad cercana a 4, siendo los últimos 5 valores los que más se alejan del valor real.

$\alpha$	0.050	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
$\beta_t/\beta_c$	4.028	3.884	3.746	3.641	4.078	4.677	4.504	3.923	4.139	3.975	3.812	3.677	3.566	3.484	3.403	3.342	3.293	3.250	3.212
$\beta_t$	-1.450	-1.441	-1.431	-1.420	-1.566	-1.754	-1.761	-1.585	-1.817	-1.880	-1.971	-2.107	-2.300	-2.582	-2.984	-3.609	-4.669	-6.808	-13.257
$\beta_c$	-0.360	-0.371	-0.382	-0.390	-0.384	-0.375	-0.391	-0.404	-0.439	-0.473	-0.517	-0.573	-0.645	-0.741	-0.877	-1.080	-1.418	-2.095	-4.127
$U_{\text{fici}}^0$	69.417	20.538	8.792	3.954	-207.882	-3.354	-2.583	-17.964	-1.415	-1.061	-0.828	-0.662	-0.542	-0.454	-0.388	-0.336	-0.295	-0.262	-0.234
$U_{\text{micro}}^0$	35.267	9.101	8.228	0.775	48.643	-1.076	-0.847	-5.016	-0.461	-0.335	-0.252	-0.192	-0.151	-0.121	-0.099	-0.082	-0.069	-0.059	-0.051
$U_{\text{auto}}^0$	30.068	8.009	1.873	-0.964	39.498	-0.378	-0.163	1.354	0.011	0.026	0.025	0.018	0.011	0.004	-0.001	-0.004	-0.007	-0.009	-0.010
$\delta$	0.000	0.000	0.000	0.000	0.002	0.123	0.107	0.004	0.108	0.118	0.128	0.140	0.154	0.168	0.181	0.194	0.207	0.219	0.231
LL	4980.116	4976.981	4974.719	4973.286	4966.023	4964.834	4963.740	4962.422	4962.106	4961.524	4961.070	4960.726	4960.479	4960.317	4960.230	4960.213	4960.259	4960.364	4960.523
TE	519.560	518.570	475.660	551.950	951.170	742.090	719.960	721.370	548.780	546.930	622.360	562.530	671.640	533.400	625.590	630.450	549.660	422.500	498.450

Tabla 4.10: Evaluación en cambios de los resultados de estimación al modificar el parámetro  $\alpha$ , siendo  $\alpha = 0,60$  el valor real con el que se generaron los datos

#### 4.2.4.4. Puntos iniciales heterogéneos

Durante toda la estimación se ha asumido que la opinión inicial (utilidad cero) es general en la población para cada una de las alternativas. Esto es poco realista, ya que es lógico pensar que todos los individuos tienen diferentes opiniones y percepciones sobre distintas alternativas en un contexto de elección.

Se seguirá realizando la estimación con los parámetros de  $U_i^0$  generales para la población, pero en esta sub-sección, se generarán los datos aleatoriamente siguiendo a una distribución uniforme. Se espera que la estimación pueda recuperar, al menos, la diferencia entre las medias de los valores de la distribución.

Las opiniones iniciales para la alternativa bicicleta va de -3 a 1 con media -1.00, la de metro-tren de -2 a 2 con media 0, micro-bus -2.5 a 1.5 con media -0.52, y finalmente la del automóvil de -2.5 a 1.5 con media -0.51.

Luego, los parámetros estimados con parámetro  $\alpha$  conocido y desconocido se observan en la

Modelo	DEoptim		BFGS	
	est.	desv.	est.	desv.
$\beta_t$	-1.490	N.R.	-1.479	0.152
$\beta_c$	-0.413	N.R.	-0.414	0.049
$U_{bici}^0$	-6.175	N.R.	-6.698	8.410
$U_{micro}^0$	-5.066	N.R.	-5.553	7.147
$U_{auto}^0$	-6.310	N.R.	-6.824	9.470
$\delta$	0.013	N.R.	0.011	0.019
$\alpha$	0.472	N.R.	0.470	0.081
LL	4987.435		4987.422	
Tiempo	180m 4s		16m 12s	

Tabla 4.11: Resultados estimación del modelo prototipo para datos generados con puntos iniciales heterogeneos

Tabla 4.11. Se puede observar que las diferencias entre los valores de las opiniones iniciales no son iguales a la media de los valores reales, y tampoco se recuperan la relación entre ellas. Sin embargo, sí se puede recuperar parte de la relación de los parámetros entre  $\beta_t$  y  $\beta_c$  (4). Además, se aprecia que para los tres parámetros de opiniones iniciales, se tiene que su desviación estándar es muy alta, comparada con el valor estimado del parámetro. Por lo tanto, es de vital importancia incluir mejoras a futuro que puedan hacerse cargo de esta problemática, para que este modelo esté completamente corregido. Además, el estimador

convexo (BFGS) estima el parámetro  $\alpha$  y el resto de los parámetros de forma similar al optimizador global DEoptim.

#### 4.2.5. Decisión de especificación del prototipo

Finalmente, para comparar con otros modelos se decide por optar la especificación del modelo que estima el parámetro  $\alpha$  y utiliza los algoritmos evolucionarios para la optimización (DEoptim). Esto, ya que el proceso de estimar con  $\alpha$  fijo, es un proceso bastante tedioso, en el cual se debe probar con algunos valores del parámetro, lo que lo vuelve poco práctico. En cambio, al estimar con optimizador global, el proceso de configuración se vuelve menos complejo. A pesar de que el método escogido toma un tiempo considerable, es el método que cumple con los requisitos para realizar una estimación transparente y que pueda capturar los valores (en cierta medida) de los parámetros.

$T = 7, \alpha = 0,6$	Valor real	BFGS	L-BFGS-B	Nelder-Mead	Malschains	DEoptim
$\beta_t/\beta_c$	4	3.677	3.673	6.019	3.675	3.675
$\beta_t$	-2.00	-2.107	-2.105	-1.890	-2.106	-2.106
$\beta_c$	-0.50	-0.573	-0.573	-0.314	-0.573	-0.573
$U_{bici}^0$	-1.00	-0.661	-0.664	1.927	-0.662	-0.662
$U_{micro}^0$	-0.5	-0.192	-0.193	1.361	-0.192	-0.192
$U_{auto}^0$	-0.5	0.018	0.019	-0.132	0.018	0.018
$\delta$	0.10	0.140	0.140	0.000	0.140	0.140
LL	—	4960.726	4960.726	4997.000	4960.726	4960.726
Tiempo	—	13m 4s	5m 56s	7m 57s	72m 5s	80m 3s

Tabla 4.12: Comparación métodos de estimación con parámetro  $\alpha$  conocido, con cantidad de pasos máxima igual a 7

De la Tabla 4.12, se puede apreciar que al tener como conocido y fijo el parámetro  $\alpha$ , todos los métodos de optimización llegan al mismo resultado, exceptuando al método Nelder-Mead. Además, el método L-BFGS-B, es el método que menos tiempo toma. Por lo tanto, al considerar  $\alpha$  fijo, se sugiere utilizar este último método para la estimación.

Al estimar con  $\alpha$  desconocido el problema cambia, ya que, el único método capaz de capturar un valor cercano del parámetro  $\alpha$  es DEoptim. 3 de los 4 métodos capturan de manera razonable el valor real de la relación entre los parámetros de los atributos. A pesar de que el valor de la verosimilitud del método DEoptim es mayor y se demora considerablemente más, se escoge utilizar este método, ya que es el que representa de mejor manera los parámetros de la simulación. Se debe estudiar a futuro, encontrar un algoritmo de optimización global

Método $T_{max} = 7$	Valor real	BFGS	LFGBS	NelderMead	Malschains (sw)	DEoptim
$\beta_t/\beta_c$	4.00	3.264	3.359	3.523	4.431	3.302
$\beta_t$	-2.00	-7.161	-3.423	-1.409	-1.575	-1.932
$\beta_c$	-0.50	-2.194	-1.019	-0.400	-0.436	-0.585
$U_{bici}^0$	-1.00	-0.257	-0.348	0.53	-4.517	-0.892
$U_{micro}^0$	-0.5	-0.057	-0.086	-0.44	0.289	-0.239
$U_{auto}^0$	-0.5	-0.009	-0.003	-0.43	1.758	-0.030
$\delta$	0.10	0.224	0.19	0.000	0.012	0.077
$\alpha$	0.60	0.904	0.787	0.27	0.471	0.622
LL	—	4960.376	4960.211	4972.368	4962.693	4960.975
Tiempo	—	11m 28s	7m 20s	7m 58s	41m 56s	86min 16s

Tabla 4.13: Comparación métodos de estimación con parámetro  $\alpha$  estimado, con cantidad de pasos máxima igual a 7

que tome menos tiempo que DEoptim y que pueda obtener resultados similares.

#### 4.2.6. Recuperación de parámetros

Elegido ya el método de optimización que se utilizará para el modelo, se propone testear si el modelo recupera exitosamente los parámetros. El optimizador DEoptim no reporta desviación estándar, por lo que se debe utilizar el optimizador BFGS con el parámetro  $\alpha$  de DEoptim como valor conocido, para poder obtener el hessiano del problema. Es importante destacar que esto se realiza solamente con el parámetro  $\alpha$  fijo, y no con todos los parámetros de DEoptim como puntos iniciales, porque al estimarlo de esta manera, los resultados difieren bastante del resultado de DEoptim como se observa en la Tabla 4.14. En donde los parámetros con puntos iniciales de DEoptim vuelven a caer en un óptimo local y el parámetro  $\alpha$  se aleja del real. En cambio, al calcularlo con  $\alpha$  como dato, se llegan a resultados similares a los del optimizador global.

	DEoptim sin hessiano		BFGS puntos iniciales DEoptim		BFGS $\alpha=0.622^*$ fijo Deoptim	
	Est.	Des. std.	Est.	Des. std.	Est.	Des. std.
	$\beta_t$	-1.932	N.R.	-3.429	2.568	-2.191
$\beta_c$	-0.585	N.R.	-1.022	0.810	-0.602	0.069
$U_{bici}^0$	-0.892	N.R.	-0.348	0.205	-0.600	0.258
$U_{micro}^0$	-0.239	N.R.	-0.086	0.104	-0.171	0.175
$U_{auto}^0$	-0.030	N.R.	-0.003	0.094	0.016	0.176
$\delta$	0.077	N.R.	0.191	0.111	0.148	0.110
$\alpha$	0.622	N.R.	0.788	0.178	—	—

Tabla 4.14: Comparación métodos finales

Luego, los resultados de la doble optimización se encuentra en la Tabla 4.15. En donde se puede observar el valor del parámetro, la desviación estándar, el test de significancia diferente de 0, y el test de igualdad al valor real para cada parámetro. Todos los parámetros a excepción de 2 de las 3 utilidades iniciales son estadísticamente iguales a su valor real (95% de confianza). Además estos 2 parámetros tampoco son estadísticamente significativos, en donde también, se considera al parámetro dentro  $\delta$  de esta categoría.

A pesar de problemas de recuperación de algunos parámetros, se puede concluir que, al menos, los  $\beta$ s de los atributos se recuperan satisfactoriamente.

	Estim	Std. Error	p-valor	p-valor valor real
$\beta_t$	-2.191	0.423	0.000	0.651
$\beta_c$	-0.602	0.069	0.000	0.138
$U_{bici}^0$	-0.600	0.258	0.020	0.122
$U_{micro}^0$	-0.171	0.175	0.330	0.061
$U_{auto}^0$	0.016	0.176	0.928	0.003
$\delta$	0.148	0.110	0.179	0.599
$\alpha^*$	0.622	N.R.	N.R.	N.R.

Tabla 4.15: Resultados de la estimación de los parámetros con el modelo prototipo

#### 4.2.7. Comparación con otros modelos

Finalmente, en esta sub-sección, se comparan diferentes modelos de estimación de los datos. Esto con el objetivo de analizar si la consideración del proceso cognitivo de la elección entrega alguna mejora en el modelamiento de elecciones.

Modelo	Mejor especificación	Mejor especificación	MNL	Mixed Logit
	con $\alpha$ fijo $T_{max} = 7$ (L-BFGS-B)	con $\alpha$ estimado $T_{max} = 7$ (DEoptim 300 it)		
$\beta_t/\beta_c$	3.677	3.302	5.27	4.56
LL	-4960.726	-4960.975	-5012.8	-5004.9
$\rho^2$	0.106	0.105	0.085	0.087
AIC	9935.4	9936.0	10035.59	10025.8
Tiempo	5m 56s	86min 16s	0s	43m 47s

Tabla 4.16: Comparación modelos con datos creados con el modelo propuesto

Al generar los datos con el prototipo, se descubre que el método de estimación que mejor ajuste es el mismo prototipo, comparándolo con MNL y Mixed Logit. Por lo tanto, al considerar los supuestos de este modelo, se logra llegar a un mejor ajuste, incluso al considerar las diferencias en grados de libertad (número de parámetros), al comparar el  $\rho^2$  y el AIC. Sin embargo, la relación entre los parámetros  $\beta_c$  y  $\beta_t$  es mejor recuperada por Mixed Logit, que por el modelo asumiendo no convexidad.

En futuras iteraciones del modelo, se debe realizar un análisis de predicción, para abarcar todas las aristas de comparación.

### 4.3. Estimación base de datos real: caso SwissMetro

En esta sección se compara el modelo de la misma forma que en la sección anterior, pero con una base de datos real. La base de datos utilizada es la de la encuesta PD/PR SwissMetro. SwissMetro es un sistema propuesto de metro-trenes que conectaría 5 de los grandes centros urbanos de Suiza. Para el estudio de pre-factibilidad de este proyecto se realizó una considerable encuesta de preferencias declaradas y preferencias reveladas (Abay, 1999; Bierlaire et al., 2001). Se encuestó a 1105 individuos entregándoles 9 situaciones ficticias de elección, en donde debían escoger entre 3 opciones: tren, SwissMetro o automóvil.

Las variables relevantes para este estudio son: tiempos de viaje y el costo que involucra este. Además, se tiene información respecto a lo respondido por los individuos en cada situación de elección.

Se realizó una estimación del modelo filtrando la base de datos por propósito de viaje, dejando solamente los propósitos 'estudios' y 'trabajo', además se filtró a los usuarios que respondieron que no tenían disponibilidad de automóvil dejando la muestra en 5607 elecciones.

Para testear el modelo prototipo se utilizaron 3 métodos para observar el comportamiento de la estimación, estos métodos de optimización son: BFGS, L-BFGS-B y DEoptim. Como se ha visto durante este informe, los dos primeros métodos toman menos tiempo de procesamiento pero tienden a caer en óptimos locales, lo cual, debido al parámetro  $\alpha$  es un problema. Se utilizan estos métodos para poder comparar qué tanto cambia respecto al optimizador global en esta base de datos. Además, es importante destacar que el optimizador global DEoptim, no entrega matriz de varianza-covarianza, por lo que no se puede obtener el valor de la desviación estándar de los parámetros (N.R. = no reporta), y por lo tanto, no se puede realizar test t. Se procede a utilizar el método DEoptim, y luego un método convexo seteando el  $\alpha$  inicial igual a al obtenido con el optimizador DEoptim.

Para el cálculo del modelo prototipo, se hizo el supuesto que el tiempo máximo de elección es de 7 pasos, esto debido al costo computacional que implica un  $T_{max}$  mayor.

Método $T_{max} = 7$	BFGS		L-BFGS-B		DEoptim 500 it		L-BFGS-B con $\alpha$ DEoptim		BFGS con $\alpha$ DEoptim	
	Media	p valor	Media	p valor	Media	p valor	valor	p.valor	Media	p valor
$\beta_t$	-2.994	0.000	-3.748	0.006	-3.241	N.R.	-3.166	0.000	-3.166	0.000
$\beta_c$	-5.146	0.000	-5.836	0.006	-5.027	N.R.	-4.928	0.000	-4.928	0.000
$U_{train}^0$	-1.239	0.000	-1.220	0.000	-1.227	N.R.	-1.238	0.000	-1.238	0.000
$U_{car}^0$	-0.152	0.000	-0.175	0.000	-0.171	N.R.	-0.175	0.000	-0.175	0.000
$\delta$	0.033	0.000	-0.032	0.000	0.035	N.R.	0.032	0.000	0.032	0.000
$\alpha$	0.979	0.000	0.983	0.000	0.979	N.R.	—	—	—	—
LL	4226.653		4227.433		4227.024		4226.944		4226.944	
Tiempo	20m 50s		13m 42s		156m 37s		156m 37s + 4m 17s		156m 37s + 6m 10s	

Tabla 4.17: Prueba de estimación del modelo prototipo con base de datos de Swissmetro mediante diferentes métodos de estimación

Es importante destacar que a pesar de que los métodos de las primeras dos columnas dan cercanos al valor que entrega el optimizador global, estos no pueden ser considerados correctos sin revisar los resultados de DEoptim. Esto debido a que, como se ha dicho durante este documento, existe no convexidad con el factor  $\alpha$ , por lo tanto, los puntos iniciales tienen influencia en los resultados que se producen para estos 2 optimizadores.

Luego, se comparan estos resultados con con otros modelos (C RRM,  $\mu$  RRM, P RRM, RAM) calculados por Belgiawan et al. (2017) siguiendo el mismo filtro de la base de datos, y, además, se agregan otros dos modelos convencionales (MNL y Mixed Logit). En Mixed Logit, se asume que los casos de elección (9 por cada individuo) son diferentes individuos, para que pueda ser directamente comparable con el modelo prototipo, el cual no hace diferenciación heterogénea para la población. El modelo Mixed Logit fue calculado utilizando el software



Biogeme (Bierlaire, 2018), en donde sólo se consideró variabilidad en los parámetros  $\beta_c$  y  $\beta_t$ .

	Prototipo		MNL		Mixed Logit		C RRM		u RRM		P RRM		RAM	
	est	t-stat	est	t-stat	est	t-stat	est	t-stat	est	t-stat	est	t-stat	est	t-stat
$\beta_t$	-3.166	-12.03	-1.27	-20.896	-2.80	-13.56	0.01	-17.8	0.01	-9.3	0.01	-8.4	0.08	-9.3
$\beta_c$	-4.928	-13.82	-1.15	-21.731	-2.28	-12.77	0.01	-16.8	0.01	-16.6	0.01	-16.2	0.08	-10.0
LL	-4226.994		-4382.5		-4211.062		-4539.672		-4373.356		-4418.252		-4239.245	
AIC	1.5099		1.5646		1.5049		1.6207		1.5617		1.5774		1.5136	
BIC	1.5166		1.5694		1.5074		1.6254		1.5677		1.5821		1.5183	
Tiempo	160m 54s		0s		2m 7s		N.R		N.R		N.R		N.R	

Tabla 4.18: Comparación modelos para la estimación de Swissmetro ( $N = 5607$ ), ultimas 4 columnas son calculadas por Belgiawan et al. (2017), escaladas en  $\times 0,01$

De la Tabla 4.18 se puede observar que el modelo obtiene mejores resultados que la mayoría de los modelos, en términos de verosimilitud y los criterios de información (Bayesiano y Akaike). Sin embargo, el prototipo propuesto tiene un tiempo de estimación considerablemente mayor al resto de los modelos, debido a que se deben utilizar optimizadores globales que toman mayor tiempo. Solamente el modelo Mixed Logit entrega mejores resultados, lo que es esperable, debido a que este modelo considera heterogeneidad en la población. Además, se observa que la relación entre parámetros del resto de los modelos ( $\beta_t/\beta_c$ ) es mayor a 1, pero en el modelo propuesto esta relación es menor a 1, lo que sugiere que quizás no se está atrapando correctamente la relación entre estas variables, o que, esta relación, tiene otro significado en el modelo. Finalmente, para analizar si efectivamente este modelo puede representar mejor las elecciones de los individuos en la base de datos de SwissMetro, se deben realizar análisis de predicción, que no serán realizados en esta memoria.

### 4.3.1. Comparación con DFT

En 2019, se publicó un software de libre acceso llamado Apollo (Hess y Palma, 2019a,b), el cual tiene base en R y en C++. Mediante este software se puede estimar el modelo Decision Field Theory, utilizando la última versión propuesta por Hancock et al. (2018).

Luego, los resultados de los parámetros del modelo Decision Field Theory para la base de datos de Swissmetro se encuentra en la Tabla 4.19. Sin embargo, para que la base de datos SwissMetro pueda ser utilizada y el modelo DFT pueda ser estimado, se tuvo que modificar el modelo, ya que DFT no acepta variables con mismos valores, es decir, si dos o más alternativas tienen los mismos valores de atributos no se pueden invertir matrices en su cálculo. Luego, se fijó en 1 la variable del error  $\varepsilon$ , y los parámetros de sensibilidad y

memoria se fijaron en 0, ya que a veces estos parámetros son insignificantes<sup>1</sup> y el realizar esta modificación arregla el problema matricial.

DFT	Est.	Err.std.	t.ratio
$ASC_{train}$	-1.120	0.107	-10.5
$ASC_{car}$	-0.076	0.074	-1.03
$\beta_t$	-1.251	0.109	-11.5
$\beta_c$	-0.867	0.064	-13.6
$p_{timesteps}$	0.564	0.178	3.16

Tabla 4.19: Resultados modelo DFT para SwissMetro

En términos generales, la comparación entre DFT y el prototipo se encuentran en la Tabla 4.20, en donde se puede apreciar que el modelo propuesto tiene mejor desempeño que DFT para estimar la verosimilitud de la base de datos de SwissMetro. Sin embargo, el prototipo toma considerablemente mucho más tiempo, por razones ya explicadas anteriormente. Además, debido a que esta es una de las primeras pruebas del modelo, puede que éste esté ajustando mejor debido a razones no consideradas en esta memoria. A pesar de lo anterior,

Parámetros generales	DFT	Prototipo
LL	-4277.17	-4226.994
$\rho^2$	0.3056	0.3137
$\rho_{adj}^2$	0.3048	0.3126
AIC/N	1.527	1.510
BIC/N	1.530	1.517
Tiempo	7m 0s	160m 54s
Nr. param.	5	6

Tabla 4.20: Comparación parámetros generales estimación SwissMetro entre Decision Field Theory y prototipo modelo

esto indica que el modelo, para la estimación de SwissMetro, y considerando solamente 2 variables (tiempo y costo) tiene un mejor desempeño, dando en cuenta que no se agregó a este modelo la heterogeneidad de los individuos, o el problema de nominalidad mencionado en 4.1.3. Es necesario probar, en futuras iteraciones, el desempeño del modelo en cuánto a predicción, ya que, en esta faceta se puede comprobar si el modelo es efectivamente mejor que DFT.

<sup>1</sup>Thomas Hancock, comunicación personal via e-mail. Respuesta pública en foro de usuarios de software Apollo: <https://groups.google.com/d/msg/apollo-choice-modelling/ATi4cFWgvr4/ZkSyx9IOBQAJ>. 13 de Septiembre 2019.

### 4.3.2. Comportamiento modelo propuesto

Con los parámetros obtenidos con el método DEoptim en la sección anterior se procede a simular el comportamiento de los individuos para observar si la evolución de utilidad es lo esperado con estos casos reales.

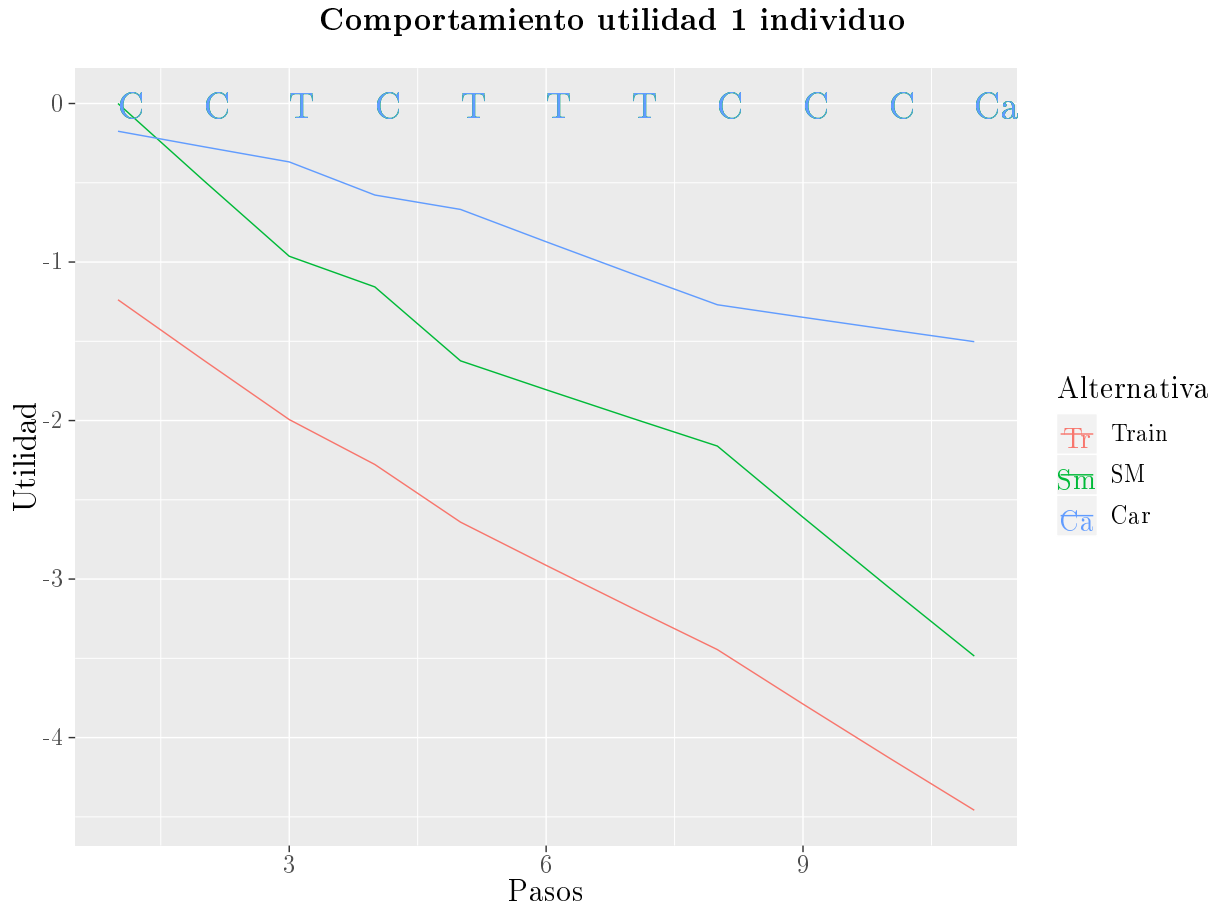


Figura 4.12: Simulación comportamiento de la utilidad de un individuo con parámetros estimados (Swissmetro), con atributos observados en cada paso: T= tiempo de viaje y C= costo de viaje

En las Figura 4.12 se observa la utilidad de uno de los individuos de la muestra, en donde se ve que la evolución de la utilidad solo decrece, y que, la diferencia entre las utilidades se va manteniendo durante el tiempo, además, no existe el fenómeno de cambio de preferencias, es decir, que una utilidad supere a otra debido al proceso de deliberación. Este comportamiento no es aislado solamente para este individuo, ya que ocurre para todos los individuos (Figura 4.13), por lo que el modelo no está reproduciendo el comportamiento esperado con los parámetros estimados.

### Comportamiento utilidad de una alternativa 40 individuos

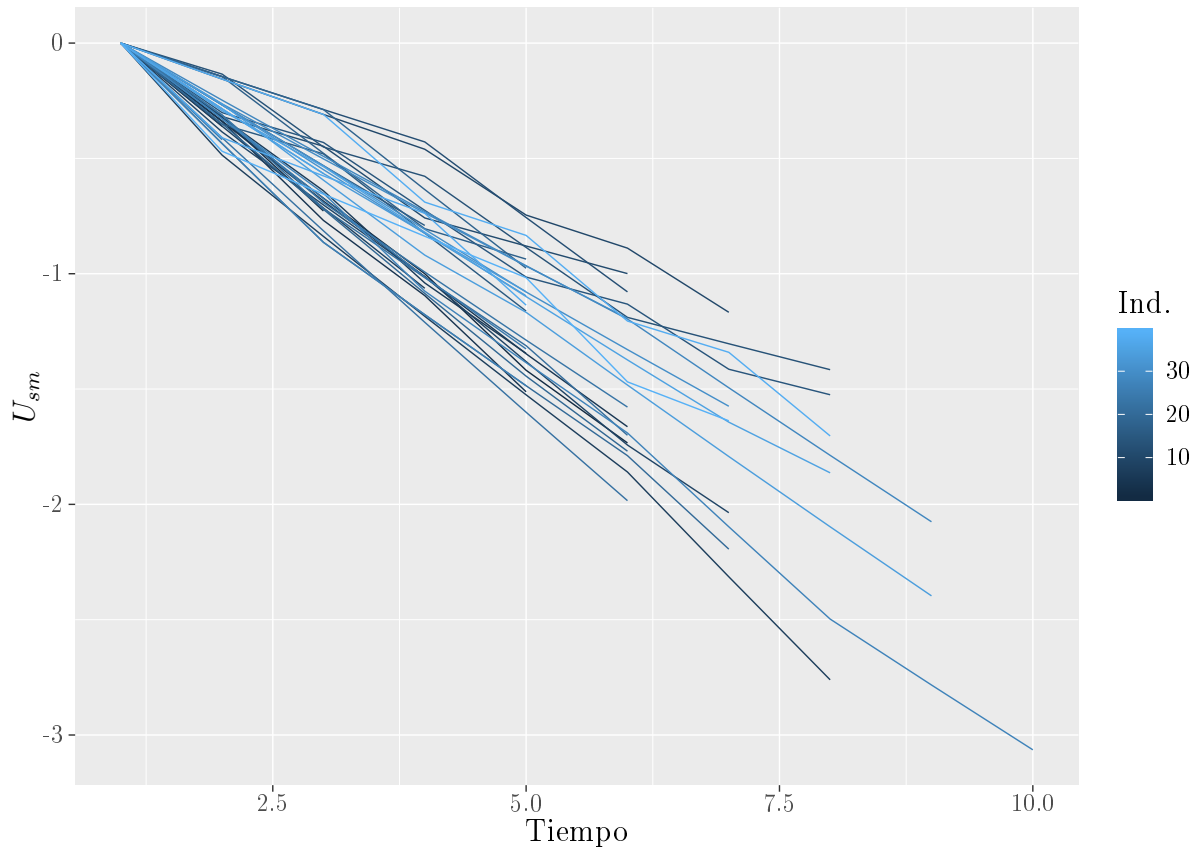


Figura 4.13: Simulación comportamiento de la utilidad de una alternativa de 40 individuos con parámetros estimados (Swissmetro)

# Capítulo 5

## Conclusiones

En esta memoria se presentó el prototipo de un modelo de elección discreta que considera el proceso cognitivo que los individuos realizan al analizar las alternativas para su elección. Este proceso de decisión se considera como información latente, por lo tanto en la formulación del prototipo se desarrolla una integración de todos los posibles casos de análisis que el individuo podría realizar. Se propuso un método de generación de datos y de estimación, y finalmente se comprobó la validez del comportamiento del modelo y su estimación.

Se testeó la validez del modelo mediante una simulación de Monte Carlo suponiendo 3 casos: que se conoce lo observado por el individuo, que se conoce el tiempo de observación y que no se tiene la información anterior. El primer caso supuso una recuperación de parámetros exitosa. Para el segundo caso, se pudo recuperar el valor de los parámetros en la gran mayoría de las simulaciones. Para el tercer caso, en donde no se conoce información extra a la normal, se pudo recuperar el valor de la mitad de los parámetros en cuestión, sin embargo, se debió utilizar un doble optimizador, por lo que futuros trabajos deben resolver la problemática de optimización.

Si bien, el modelo precisa de ajustes que se mencionarán en este capítulo, se llegó a obtener resultados positivos respecto a su estimación, al relajar algunos de los supuestos de éste (por ejemplo, fijar los parámetros de escala). Al generar los datos suponiendo que los individuos se comportan de acuerdo a los supuestos del modelo, se pudo observar que al estimar los parámetros, el prototipo ajusta mejor que otros otros métodos de la literatura (Mixed Logit y MNL).

Al estimar el modelo utilizando una base de datos real (SwissMetro), se obtuvieron mejores ajustes que modelos que se consideran en la literatura, lo que indica que el prototipo, dadas ciertas condiciones, puede lograr una explicación de los fenómenos que ocurren en los

contextos de elecciones discretas, al considerar sólo 2 atributos en su cálculo. Solamente el modelo Mixed Logit fue superior en las métricas de desempeño.

Para la misma base de datos, se estimó mediante el modelo Decision Field Theory y se realizó una comparación con el prototipo, en donde el prototipo obtuvo un mejor ajuste de estimación que DFT. Sin embargo, no existe claridad aún si este mejor ajuste se debe a que los supuestos realizados en esta memoria son correctos, o se desempeña mejor por otras razones no consideradas o no conocidas, ya que al tratar de replicar el comportamiento de los individuos con el prototipo, se observa que éste es poco congruente con lo que se espera. Además, no se realizaron análisis de predicción, ya que requiere de mayor profundización que no iba acorde a los alcances de esta memoria, lo que debe ser considerado a futuro. Sin embargo, sí se están logrando, al menos, buen desempeño en términos de las medidas comparativas tradicionales (AIC, BIC, LL,  $\rho^2$ ).

No obstante, y por el hecho de que esta propuesta es un prototipo de un futuro modelo, se requieren varios ajustes y consideraciones para su mejora en trabajos futuros.

Una de las grandes limitaciones del modelo propuesto, es que debido a que se intenta proponer un método que requiera solamente la utilización de un software y no sobre-complicar su configuración, el costo computacional para estimar el modelo es bastante alto. Esto ya que se debe integrar respecto a un espacio no menor de caminos de observación de los individuos que aumenta considerablemente ( $K^t$ ) por cada paso de observación extra que se quiera agregar. Sin embargo, para estimación de espacios pequeños el modelo logra ajustarse a los datos asumiendo que sus elecciones se realizaron en pocos pasos.

Dada a la falencia anterior, es decir, que no se puede estimar gran cantidad de pasos, y agregando que los pasos son periodos de tiempo pre-definidos, el modelo no considera cuánto es el tiempo real del proceso de deliberación por el que pasan los individuos. Es decir, no se sabe con exactitud qué porcentaje del proceso de deliberación se abarca en la modelación. Por lo tanto, es de vital importancia realizar experimentos mediante tecnología eyetracking (u otra) para poder tener claridad respecto a qué observan los individuos, cuánto tiempo les toma realizar análisis de atributos y finalmente, cuánto dura el proceso de deliberación en elecciones de corto plazo.

Respecto a su formulación, el modelo tiene problemas que se intentaron solucionar, en cierta medida, al estudiarlo en esta memoria. El primer problema es que algunos parámetros ( $\alpha$ ,  $\mu_e$  y  $\mu_s$ ) generan que la función de verosimilitud se vuelva no convexa, por lo tanto, se requiere de optimizadores globales que requieren un tiempo mucho mayor para poder lograr la estimación de los parámetros, lo que complejiza aún más el problema de costo

computacional. Por lo tanto, para no abarcar tanto tiempo en la estimación del modelo, se impuso que  $\mu = \mu_e = \mu_s$ .

Además, dada la formulación del modelo, se descubrió durante la experimentación que si existen parámetros que tienen menor efecto negativo o parámetros positivos, éstos tendrán más posibilidad de ser observados, lo que no ocurre en el supuesto que se realizó para formular este modelo. La posibilidad de observar un atributo en un paso  $t$  debería depender de la importancia que tiene el atributo para los individuos, no de si entrega menos o más utilidad. Posibles soluciones a esta problemática es considerar que  $\mathbb{P}(k(t))$  no dependa de la utilidad posible que entregaría el atributo observado, sino que dependa del impacto que tiene el atributo en la utilidad, es decir, que dependa de  $|\beta_k * X_k|$ . No se propone que dependa solamente del valor absoluto de  $\beta_k$  ya que este puede cambiar de acuerdo a la escala que tenga el valor  $X_k$ . Otras propuestas deben ser consideradas.

Para el trabajo futuro es importante considerar la heterogeneidad de las utilidades iniciales, ya que se puede explicar de mejor manera el modelo, debido a que es claro que cada individuo tiene una opinión diferente preconcebida de cada alternativa. Esto cambiaría positivamente el modelo, ya que no se estaría generalizando un parámetro relevante para la explicación del comportamiento de los individuos de acuerdo a este modelo. Al agregar parámetros heterogéneos para la población, se espera que el modelo pueda tener un aun mejor valor del likelihood, el único problema es que los tiempos de estimación aumentarán considerablemente, asumiendo que se mantiene todo el resto del modelo igual.

Además, es de vital importancia encontrar métodos que disminuyan el tiempo de estimación del modelo, ya que con los tiempos actuales que toma calcularlo, no se podría considerar a este prototipo como una opción a los modelos tradicionales.

Futuros ajustes se pueden realizar que consideren: mejorar el problema matricial reduciendo el espacio de caminos posibles, realizar un método de estimación secuencial para los parámetros  $u$ , encontrar el mejor optimizador global para el modelo, cambiar probabilidad de observar atributo  $k$ , suavizar el problema de puntos iniciales trasladados incluyendo  $1/\alpha$  en la utilidad futura, formular un método de predicción y finalmente, incorporar heterogeneidad a las opiniones iniciales. De ser necesario, se puede replantear la función de utilidad, que mantenga la lógica planteada en esta memoria, pero que pueda hacerse cargo de algunas limitaciones.

Finalmente, se concluye, que el prototipo requiere de ajustes y mejoras, pero sienta las bases a un modelo que puede ser una alternativa a los modelos convencionales, ya que en su primera iteración ya entrega resultados prometedores respecto a las medidas tradicionales de

desempeño del ajuste de estimación.



# Apéndice A

## Anexo: Código R ejemplo generación de datos

```
#Generacion de elecciones
N<-4000 #individuos
#Tiempos de viaje:
TVbici  <-abs(runif(N, 10, 20))/10 #Bicicleta
TVauto  <-abs(runif(N, 2, 25))/10 #Automovil
TVmetro <-abs(runif(N, 4, 15))/10 #Metro-tren
TVmicro <-abs(runif(N, 3, 22))/10 #Micro-bus
#80% paga mayor tarifa que 20%:
poblacion <- rbinom(N, 1, 0.8)
#Precios de viaje:
COSTmicro<-c()
COSTmetro<-c()
COSTmicro[poblacion==1]<-700/100
COSTmicro[poblacion==0]<-230/100
COSTmetro[poblacion==1]<-800/100
COSTmetro[poblacion==0]<-230/100
COSTbici  <- runif(N, 0, 500)/100
COSTauto  <- runif(N, 0, 1800)/100
ind <- 1:N #Lista individuos
#Matriz data:
mat <- cbind(ind, TVbici,TVmicro, TVmetro, TVauto,
COSTbici,COSTmicro, COSTmetro, COSTauto)
#Vector de unos
ones<-rep(1,N)
```

```

#Matrices diferenciadas por atributo
TDV <-cbind(TVbici ,TVmicro ,TVmetro ,TVauto)
COST<-cbind(COSTbici ,COSTmicro ,COSTmetro ,COSTauto)
#Matrices diferenciadas por alternativa
bici  <-cbind(TVbici ,COSTbici)
micro <-cbind(TVmicro ,COSTmicro)
metro <-cbind(TVmetro ,COSTmetro)
auto  <-cbind(TVauto ,COSTauto)
#Puntos iniciales homogeneos
OIbici  <--rep(1,N)
OImetro <--rep(0,N)
OImicro <--rep(0.5,N)
OIauto  <--rep(0.5,N)
#Parametros
Ft      <-  0.6  #Alfa
Bt      <-  -2   #Beta tiempo
Bc      <- -0.5  #Beta costo
TTmax   <-  25   #Tiempo maximo de eleccion
desct   <-0.09  #Delta
escalae<-0.01  #u_e
escalas<-0.01  #u_s
escala  <-0.01  #u
MFINAL<-matrix(ncol=7) #Matriz con resultados
#Para cada individuo:
for (n in 1:N) {
  TDVn<-TDV[n,]      #Vector atributo 1 ind. n
  COSTn<-COST[n,]   #Vector atributo 2 ind. n
  #Opiniones iniciales
  Oibici  <-OIbici[n]
  Oimicro <-OImicro[n]
  Oimetro <-OImetro[n]
  Oiauto  <-OIauto[n]
  #Crear vector utilidad por individuo
  Ubici   <-vector(length=TTmax)
  Umicro  <-vector(length=TTmax)
  Umetro  <-vector(length=TTmax)
  Uauto   <-vector(length=TTmax)
  #Vector que contiene el camino tomado por el individuo
  At <- vector(length=TTmax)
  #Se inicia el proceso de deliberacion para el individuo

```

```

#En R t=1 es el t=0 real
for (t in 1:TTmax){
  desc <- desc*(t-1)^2 #Delta_t con el tiempo
  #Para t=0
  if (t==1){ #Para t=0
    #Utilidades iniciales
    Ubici[t] <- Oibici
    Umicro[t] <- Oimicro
    Umetro[t] <- Oimetro
    Uauto[t] <- Oiauto
    #Utilidades futuras si observo atributo costo
    UFCbici <- Ft*Ubici[t] +(1-Ft)*Bc*COSTn[1]
      + rgumbel(1, loc=0, scale=escalae)
    UFCmicro <- Ft*Umicro[t] +(1-Ft)*Bc*COSTn[2]
      + rgumbel(1, loc=0, scale=escalae)
    UFCmetro <- Ft*Umetro[t] +(1-Ft)*Bc*COSTn[3]
      + rgumbel(1, loc=0, scale=escalae)
    UFCauto <- Ft*Uauto[t] +(1-Ft)*Bc*COSTn[4]
      + rgumbel(1, loc=0, scale=escalae)
    #Utilidades futuras si observo atributo tiempo de viaje
    UFTbici <- Ft*Ubici[t] +(1-Ft)*Bt*TDVn[1]
      + rgumbel(1, loc=0, scale=escalae)
    UFTmicro <- Ft*Umicro[t] +(1-Ft)*Bt*TDVn[2]
      + rgumbel(1, loc=0, scale=escalae)
    UFTmetro <- Ft*Umetro[t] +(1-Ft)*Bt*TDVn[3]
      + rgumbel(1, loc=0, scale=escalae)
    UFTauto <- Ft*Uauto[t] +(1-Ft)*Bt*TDVn[4]
      + rgumbel(1, loc=0, scale=escalae)
    #Esperanza maximo de ut. futuras de costo y precio
    UFC <- (1/ue)*log((exp(ue*UFCbici) + exp(ue*UFCmicro)
      + exp(ue*UFCmetro) + exp(ue*UFCauto)))
      + rgumbel(1, loc=0, scale=escalas)
    UFT <- (1/ue)*log((exp(ue*UFTbici) + exp(ue*UFTmicro)
      + exp(ue*UFTmetro) + exp(ue*UFTauto)))
      + rgumbel(1, loc=0, scale=escalas)
    #Utilidad de seguir observando
    VS <- (1/us)*log(exp(us*UFC) + exp(us*UFT))
      + rgumbel(1, loc=0, scale=escala)
    #Utilidad de elegir
    VE <- (1/ue)*log(exp(ue*Ubici[t]) + exp(ue*Umicro[t]))

```

```

+ exp(ue*Umetro[t]) + exp(ue*Uauto[t]))
+ rgumbel(1, loc=0, scale=escala)
#Denominadores probs.
denomS <- exp(us*UFC) + exp(us*UFT)
denomE <- exp(ue*Ubici[t]) + exp(ue*Umicro[t])
+ exp(ue*Umetro[t]) + exp(ue*Uauto[t])
denom <-exp(u*VS) + exp(u*VE)
#Probabilidad de elegir alternativa
P1 <- exp(ue*Ubici[t]) /denomE
P2 <- exp(ue*Umicro[t]) /denomE
P3 <- exp(ue*Umetro[t]) /denomE
P4 <- exp(ue*Uauto[t]) /denomE
#Probabilidad de observar atributo
PC <- exp(us*UFC)/denomS
PT <- exp(us*UFT)/denomS
#Probabilidad de seguir observando
PS <- exp(u*VS)/(exp(u*VS)+exp(u*(VE+desc)))
+ exp(u*VE)/(exp(u*(VS+desc))+exp(u*VE))
#Probabilidad de elegir
PE <- 1-PS
#Probabilidades acumuladas para Random
Pac <- cumsum(c(P1*PE, P2*PE, P3*PE, P4*PE, PC*PS, PT*PS))
Random <- runif(1, min = 0, max = 1)
if (Random<Pac[1]){At[t]<-1}#Bicicleta
else if (Random>Pac[1] & Random<Pac[2]){At[t]<-2}#Metro
else if (Random>Pac[2] & Random<Pac[3]){At[t]<-3}#Micro
else if (Random>Pac[3] & Random<Pac[4]){At[t]<-4}#Auto
else if (Random>Pac[4] & Random<Pac[5]){At[t]<-5}#Costo
else {At[t]<-6} #Precio
}
#Para t=limite
else if(t==TTmax){
if (At[t-1]==1){At[t]<-0}
else if (At[t-1]==0){At[t]<-0}
else if (At[t-1]==2){At[t]<-0}
else if (At[t-1]==3){At[t]<-0}
else if (At[t-1]==4){At[t]<-0}
else if (At[t-1]==5){ #si decidi observar costo
#Utilidad presente entonces es
Ubici[t] <- Ft*Ubici[t-1] + (1-Ft)*Bc*COSTn[1]

```

```

        + rgumbel(1, loc=0, scale=escalae)
Umicro[t] <- Ft*Umicro[t-1] + (1-Ft)*Bc*COSTn[2]
        + rgumbel(1, loc=0, scale=escalae)
Umetro[t] <- Ft*Umetro[t-1] + (1-Ft)*Bc*COSTn[3]
        + rgumbel(1, loc=0, scale=escalae)
Uauto[t] <- Ft*Uauto[t-1] + (1-Ft)*Bc*COSTn[4]
        + rgumbel(1, loc=0, scale=escalae)
#Como estoy en tiempo limite, no existe probabilidad
#de seguir observando
#Denominador
denom <- exp(ue*Ubici[t]) + exp(ue*Umicro[t])
        + exp(ue*Umetro[t]) + exp(ue*Uauto[t])
#Probabilidad de elegir alternativa
P1 <- exp(ue*Ubici[t]) / denom
P2 <- exp(ue*Umicro[t]) / denom
P3 <- exp(ue*Umetro[t]) / denom
P4 <- exp(ue*Uauto[t]) / denom
#Probabilidad acumulada ->Eleccion
Pac <- cumsum(c(P1,P2,P3,P4))
Random<- runif(1, min = 0, max = 1)
if (Random<Pac[1]) {At[t]<-1}
else if (Random>Pac[1] & Random<Pac[2]){At[t]<-2}
else if (Random>Pac[2] & Random<Pac[3]){At[t]<-3}
else {At[t]<-4}
}
else { #si decidi observar precio
#Utilidad presente entonces es
Ubici[t] <- Ft*Ubici[t-1] + (1-Ft)*Bt*TDVn[1]
        + rgumbel(1, loc=0, scale=escalae)
Umicro[t] <- Ft*Umicro[t-1] + (1-Ft)*Bt*TDVn[2]
        + rgumbel(1, loc=0, scale=escalae)
Umetro[t] <- Ft*Umetro[t-1] + (1-Ft)*Bt*TDVn[3]
        + rgumbel(1, loc=0, scale=escalae)
Uauto[t] <- Ft*Uauto[t-1] + (1-Ft)*Bt*TDVn[4]
        + rgumbel(1, loc=0, scale=escalae)
#Probabilidades
denom <- exp(ue*Ubici[t]) + exp(ue*Umicro[t])
        + exp(ue*Umetro[t]) + exp(ue*Uauto[t])
P1 <- exp(ue*Ubici[t]) / denom
P2 <- exp(ue*Umicro[t]) / denom

```

```

P3      <-  exp(ue*Umetro[t]) / denom
P4      <-  exp(ue*Uauto[t]) / denom
#Probabilidad acumulada ->Eleccion
Pac      <- cumsum(c(P1,P2,P3,P4))
Random<- runif(1, min = 0, max = 1)
if      (Random<Pac[1])                {At[t]<-1}
else if (Random>Pac[1] & Random<Pac[2]){At[t]<-2}
else if (Random>Pac[2] & Random<Pac[3]){At[t]<-3}
else                {At[t]<-4}
}
}
#tiempo distinto a 0 y limite:
else {
  if      (At[t-1]==1){At[t]<-0}
  else if (At[t-1]==0){At[t]<-0}
  else if (At[t-1]==2){At[t]<-0}
  else if (At[t-1]==3){At[t]<-0}
  else if (At[t-1]==4){At[t]<-0}
  else if (At[t-1]==5){ #si decidi observar costo
    #Utilidades presentes
    Ubici[t] <- Ft*Ubici[t-1] + (1-Ft)*Bc*COSTn[1]
                + rgumbel(1, loc=0, scale=escalae)
    Umicro[t] <- Ft*Umicro[t-1] + (1-Ft)*Bc*COSTn[2]
                + rgumbel(1, loc=0, scale=escalae)
    Umetro[t] <- Ft*Umetro[t-1] + (1-Ft)*Bc*COSTn[3]
                + rgumbel(1, loc=0, scale=escalae)
    Uauto[t] <- Ft*Uauto[t-1] + (1-Ft)*Bc*COSTn[4]
                + rgumbel(1, loc=0, scale=escalae)
    #Utilidades futuras de observar costo
    UFCbici <- Ft*Ubici[t] + (1-Ft)*Bc*COSTn[1]
                + rgumbel(1, loc=0, scale=escalae)
    UFCmicro <- Ft*Umicro[t] + (1-Ft)*Bc*COSTn[2]
                + rgumbel(1, loc=0, scale=escalae)
    UFCmetro <- Ft*Umetro[t] + (1-Ft)*Bc*COSTn[3]
                + rgumbel(1, loc=0, scale=escalae)
    UFCauto <- Ft*Uauto[t] + (1-Ft)*Bc*COSTn[4]
                + rgumbel(1, loc=0, scale=escalae)
    #Utilidades futuras de observar tiempo
    UFTbici <- Ft*Ubici[t] + (1-Ft)*Bt*TDVn[1]
                + rgumbel(1, loc=0, scale=escalae)

```

```

UFTmicro <- Ft*Umicro[t] + (1-Ft)*Bt*TDVn[2]
           + rgumbel(1, loc=0, scale=escalae)
UFTmetro <- Ft*Umetro[t] + (1-Ft)*Bt*TDVn[3]
           + rgumbel(1, loc=0, scale=escalae)
UFTauto  <- Ft*Uauto[t]  + (1-Ft)*Bt*TDVn[4]
           + rgumbel(1, loc=0, scale=escalae)
#Esperanza utilidad maxima futuras atributos
UFC <- (1/ue)*log((exp(ue*UFCbici) + exp(ue*UFCmicro)
                 + exp(ue*UFCmetro) + exp(ue*UFCauto)))
           + rgumbel(1, loc=0, scale=escalas)
UFT <- (1/ue)*log((exp(ue*UFTbici) + exp(ue*UFTmicro)
                 + exp(ue*UFTmetro) + exp(ue*UFTauto)))
           + rgumbel(1, loc=0, scale=escalas)
#Utilidad de seguir observando
VS <- (1/us)*log(exp(us*UFC) + exp(us*UFT))
           + rgumbel(1, loc=0, scale=escala)
#Utilidad de elegir
VE <- (1/ue)*log(exp(ue*Ubici[t]) + exp(ue*Umicro[t])
                 + exp(ue*Umetro[t]) + exp(ue*Uauto[t]))
           + rgumbel(1, loc=0, scale=escala)
#Probabilidades
denomS <- exp(us*UFC)+exp(us*UFT)
denomE <- exp(ue*Ubici[t]) + exp(ue*Umicro[t])
           + exp(ue*Umetro[t]) + exp(ue*Uauto[t])
denom <- exp(u*VS)+exp(u*VE)
P1 <- exp(ue*Ubici[t]) /denomE
P2 <- exp(ue*Umicro[t]) /denomE
P3 <- exp(ue*Umetro[t]) /denomE
P4 <- exp(ue*Uauto[t]) /denomE
PC <- exp(us*UFC)/denomS
PT <- exp(us*UFT)/denomS
PS <- exp(u*VS)/(exp(u*VS) + exp(u*(VE+desc)))
           + exp(u*VE)/(exp(u*(VS+desc)) + exp(u*VE))
PE <- 1-PS
#Probabilidad acumulada ->Eleccion
Pac <- cumsum(c(P1*PE, P2*PE, P3*PE, P4*PE, PC*PS, PT*PS))
Random <- runif(1, min = 0, max = 1)
if      (Random<Pac[1]){At[t]<-1}
else if (Random>Pac[1] & Random<Pac[2]){At[t]<-2}
else if (Random>Pac[2] & Random<Pac[3]){At[t]<-3}

```

```

else if (Random>Pac[3] & Random<Pac[4]){At[t]<-4}
else if (Random>Pac[4] & Random<Pac[5]){At[t]<-5}
else {At[t]<-6}
}
else{ #si decidi observar tiempo
Ubici[t] <- Ft*Ubici[t-1] + (1-Ft)*Bt*TDVn[1]
          + rgumbel(1, loc=0, scale=escalae)
Umicro[t] <- Ft*Umicro[t-1] + (1-Ft)*Bt*TDVn[2]
          + rgumbel(1, loc=0, scale=escalae)
Umetro[t] <- Ft*Umetro[t-1] + (1-Ft)*Bt*TDVn[3]
          + rgumbel(1, loc=0, scale=escalae)
Uauto[t] <- Ft*Uauto[t-1] + (1-Ft)*Bt*TDVn[4]
          + rgumbel(1, loc=0, scale=escalae)
UFCbici <- Ft*Ubici[t] + (1-Ft)*Bc*COSTn[1]
          + rgumbel(1, loc=0, scale=escalae)
UFCmicro <- Ft*Umicro[t] + (1-Ft)*Bc*COSTn[2]
          + rgumbel(1, loc=0, scale=escalae)
UFCmetro <- Ft*Umetro[t] + (1-Ft)*Bc*COSTn[3]
          + rgumbel(1, loc=0, scale=escalae)
UFCauto <- Ft*Uauto[t] + (1-Ft)*Bc*COSTn[4]
          + rgumbel(1, loc=0, scale=escalae)
UFTbici <- Ft*Ubici[t] + (1-Ft)*Bt*TDVn[1]
          + rgumbel(1, loc=0, scale=escalae)
UFTmicro <- Ft*Umicro[t] + (1-Ft)*Bt*TDVn[2]
          + rgumbel(1, loc=0, scale=escalae)
UFTmetro <- Ft*Umetro[t] + (1-Ft)*Bt*TDVn[3]
          + rgumbel(1, loc=0, scale=escalae)
UFTauto <- Ft*Uauto[t] + (1-Ft)*Bt*TDVn[4]
          + rgumbel(1, loc=0, scale=escalae)
UFC <- (1/ue)*log((exp(ue*UFCbici) + exp(ue*UFCmicro)
          + exp(ue*UFCmetro) + exp(ue*UFCauto)))
          + rgumbel(1, loc=0, scale=escalas)
UFT <- (1/ue)*log((exp(ue*UFTbici) + exp(ue*UFTmicro)
          + exp(ue*UFTmetro) + exp(ue*UFTauto)))
          + rgumbel(1, loc=0, scale=escalas)
VS <- (1/us)*log(exp(us*UFC)+exp(us*UFT))
          + rgumbel(1, loc=0, scale=escala)
VE <- (1/ue)*log(exp(ue*Ubici[t])+exp(ue*Umicro[t])
          + exp(ue*Umetro[t])+exp(ue*Uauto[t]))
          + rgumbel(1, loc=0, scale=escala)

```



```

#Probabilidades
denomS <- exp(us*UFC)+exp(us*UFT)
denomE <- exp(ue*Ubici[t]) + exp(ue*Umicro[t])
          + exp(ue*Umetro[t]) + exp(ue*Uauto[t])
P1 <- exp(ue*Ubici[t]) /denomE
P2 <- exp(ue*Umicro[t]) /denomE
P3 <- exp(ue*Umetro[t]) /denomE
P4 <- exp(ue*Uauto[t]) /denomE
PC <- exp(us*UFC)/denomS
PT <- exp(us*UFT)/denomS
PS <- exp(u*VS)/(exp(u*VS)+exp(u*(VE+desc)))
          + exp(u*VE)/(exp(u*(VS+desc))+exp(u*VE))
PE<- 1-PS
#Probabilidad acumulada ->Eleccion
Pac <- cumsum(c(P1*PE,P2*PE,P3*PE,P4*PE,PC*PS,PT*PS))
Random <- runif(1, min = 0, max = 1)
if (Random<Pac[1]){At[t]<-1}
else if (Random>Pac[1] & Random<Pac[2]){At[t]<-2}
else if (Random>Pac[2] & Random<Pac[3]){At[t]<-3}
else if (Random>Pac[3] & Random<Pac[4]){At[t]<-4}
else if (Random>Pac[4] & Random<Pac[5]){At[t]<-5}
else {At[t]<-6}
}
}
}
ind <- rep(n,TTmax)
tiempo <- 1:TTmax
Mn <- cbind(ind,tiempo,At,Ubici,Umicro,Umetro,Uauto)
MFINAL <- rbind(MFINAL,Mn)
}
MFINAL <- MFINAL[-1,]
MFINAL <- data.frame(MFINAL)
#Creacion vector elecciones
MFINAL2 <- MFINAL[!(MFINAL$At==0),]
tiempoxid <- aggregate(MFINAL2$tiempo, by = list(MFINAL2$ind), max)
eleccion <- c()
for (i in 1:N){
tfinal <- tiempoxid[i,2]
ele <- MFINAL2[MFINAL2$ind==i & MFINAL2$tiempo==tfinal,3]
eleccion<- c(eleccion,ele) }

```

# Apéndice B

## Anexo: Código R ejemplo estimación

```
#Estimacion
TMAX <- 8 #tiempo maximo de estimacion (Tmax=7)
#Funcion de log-verosimilitud
loglikel <- function(beta){
  #Parametros a estimar
  Bt      <- beta[1] #TDV
  Bc      <- beta[2] #Cost
  Oibici  <- beta[3]
  Oimicro <- beta[4]
  Oimetro <- 0
  Oiauto  <- beta[5]
  desct   <- abs(beta[6])
  alfa    <- abs(beta[7])
  #Parametros fijos
  u       <- 1
  ue      <- 1
  us      <- 1
  #Configuracion dimension de matrices
  ALT <- 2          #cantidad de alternativas
  t   <- 0:(TMAX - 1) #vector con todos los tiempos en orden
  H   <- sum(ALT ^ t) #cantidad de caminos posibles hasta Tmax
  T2  <- TMAX - 2    #Tmax-1 (recordar que en R parte de 1)
  t2  <- 0:T2        #vector tiempos en orden Tmax-1
  H2  <- sum(ALT ^ t2)#cantidad de caminos posibles hasta Tmax-1
  #Matrices con info de cada camino
  Vst  <- matrix(nrow = N, ncol = H2)#Utilidad fut. tiempo
  Vsc  <- matrix(nrow = N, ncol = H2)#Utilidad fut. costo
```

```

Vs      <- matrix(nrow = N, ncol = H2) #Utilidad futura
Ve      <- matrix(nrow = N, ncol = H2) #Utilidad presente
CAMINO  <- matrix(nrow = N, ncol = H) #Probabilidad camino
Vbici   <- matrix(nrow = N, ncol = H) #Utilidad bicicleta
Vmicro  <- matrix(nrow = N, ncol = H) #Utilidad microbus
Vmetro  <- matrix(nrow = N, ncol = H) #Utilidad metrotren
Vauto   <- matrix(nrow = N, ncol = H) #Utilidad automovil
PS      <- matrix(nrow = N, ncol = H) #Probabilidad seguir
PE      <- matrix(nrow = N, ncol = H) #Probabilidad elegir
Prbicim <- matrix(nrow = N, ncol = H) #Probabilidad bici
Prmicrom <- matrix(nrow = N, ncol = H) #Probabilidad micro
Prmetrom <- matrix(nrow = N, ncol = H) #Probabilidad metro
Prautom <- matrix(nrow = N, ncol = H) #Probabilidad auto
#Rellenar matrices con la data de acuerdo a parametros:
#Matrices de utilidades presentes
for (h in 1:H) { #para cada camino
  if (h == 1) { #si camino es t=0
    Vbici[, h] <- 0ibici
    Vmicro[, h] <- 0imicro
    Vmetro[, h] <- 0imetro
    Vauto[, h] <- 0iauto
  }
  else{#si camino es distinto al t=0
    #lo siguiente es valido solamente para procesos
    #con 2 alternativas, si se tienen mas alternativas
    #se necesita utilizar matriz de herencia
    aux <- h %% 2 #variable aux. que atrapa la raiz
    if (h %% 2 == 0) { #si es lado izquierdo del nido
      # se observa tiempo
      Vbici[, h] <- alfa * Vbici[, aux]
        + (1 - alfa) * Bt * bici[, 1]
      Vmicro[, h] <- alfa * Vmicro[, aux]
        + (1 - alfa) * Bt * micro[, 1]
      Vmetro[, h] <- alfa * Vmetro[, aux]
        + (1 - alfa) * Bt * metro[, 1]
      Vauto[, h] <- alfa * Vauto[, aux]
        + (1 - alfa) * Bt * auto[, 1]
    }
    else{#si es lado derecho del nido
      #se observa costo

```

```

Vbici[, h] <- alfa * Vbici[, aux]
            + (1 - alfa) * Bc * bici[, 2]
Vmicro[, h] <- alfa * Vmicro[, aux]
            + (1 - alfa) * Bc * micro[, 2]
Vmetro[, h] <- alfa * Vmetro[, aux]
            + (1 - alfa) * Bc * metro[, 2]
Vauto[, h] <- alfa * Vauto[, aux]
            + (1 - alfa) * Bc * auto[, 2]
}
}
}
#Matrices de utilidades futuras de atributos
for (h in 1:H2) {
  aux1 <- 2*h      #variable aux nido izquierdo; tiempo
  aux2 <- 2*h + 1 #variable aux nido derecho; costo
  Vst[, h] <- (1 / ue) * log(exp(ue * Vbici[, aux1])
    + exp(ue * Vmicro[, aux1]) + exp(ue * Vmetro[, aux1])
    + exp(ue * Vauto[, aux1]))
  Vsc[, h] <- (1 / ue) * log(exp(ue * Vbici[, aux2])
    + exp(ue * Vmicro[, aux2]) + exp(ue * Vmetro[, aux2])
    + exp(ue * Vauto[, aux2]))
}
#Matrices de utilidad futura y presente
for (h in 1:H2) {
  Vs[,h] <- (1 / us) * log(exp(us * Vst[,h]) + exp(us * Vsc[,h]))
  Ve[,h] <- (1 / ue) * log(exp(ue * Vbici[,h])
    + exp(ue * Vmicro[,h]) + exp(ue * Vmetro[,h])
    + exp(ue * Vauto[,h]))
}
#Obtener probabilidad de seguir buscando
for (h in 1:H2) {
  t      <- log2(h) %/% 1 #Obtener tiempo de camino
  desc   <- desct*(t^2)   #aplicar factor dscto por tiempo
  PS[,h] <- (exp(u*Vs[,h])/(exp(u*Vs[,h]) + exp(u*(Ve[,h] + desc))))
    + (exp(u*Ve[,h]) / (exp(u*(Vs[,h] + desc)) + exp(u*Ve[,h])))
}
#Por problemas de computacion, a veces
#se obtienen valores mayores a 1, pero
#insignificantes, como 1.000000000001:
PS[PS>1]<-1

```

```

#Probabilidades generales por camino
PE <- 1 - PS
#Rellenar ultimas columnas H-H2 (tiempo limite)
PE[is.na(PE)] <- 1
PS[is.na(PS)] <- 0
#Probabilidades de atributos por camino
PT <- exp(us*Vst) / (exp(us*Vst) + exp(us*Vsc))
PC <- exp(us*Vsc) / (exp(us*Vst) + exp(us*Vsc))
#Probabilidades alt. por camino
denom <- exp(ue*Vbici) + exp(ue*Vmicro)
          + exp(ue*Vmetro) + exp(ue*Vauto)
P1 <- exp(ue*Vbici) / denom
P2 <- exp(ue*Vmicro) / denom
P3 <- exp(ue*Vmetro) / denom
P4 <- exp(ue*Vauto) / denom
#Obtener probabilidades de caminos
for (h in 1:H) {
  aux <- h %/% 2
  h2 <- h
  camino <- 1
  while (aux > 0) {
    if (h2 %/% 2 == 0) {
      camino <- camino * PS[, aux] * PT[, aux]
      h2 <- aux
      aux <- aux %/% 2
    }
    else{
      camino <- camino * PS[, aux] * PC[, aux]
      h2 <- aux
      aux <- aux %/% 2
    }
  }
  CAMINO[,h] <- camino
}
CAMINO <- CAMINO*PE #probabilidades de caminos completa
#Probabilidad de alternativa por camino
Prbicim <- P1 * CAMINO
Prmicrom <- P2 * CAMINO
Prmetrom <- P3 * CAMINO
Prautom <- P4 * CAMINO

```

```

#Probabilidad de alternativa por individuo
Prbici <- rowSums(Prbicim)
Prmicro <- rowSums(Prmicrom)
Prmetro <- rowSums(Prmetrom)
Prauto <- rowSums(Prautom)
#Log-verosimilitud
lli <- log((Prbici * choice[, 1]) + (Prmicro * choice[, 2])
+ (Prmetro * choice[, 3])+(Prauto * choice[, 4]))
return(-sum(lli)) #Output funcion
}

#Ejemplo optimizacion BFGS
mymle0 <- optim(par=BOP, fn=loglikel, hessian=TRUE,
method='BFGS', control = list(maxit=30000, trace=TRUE, REPORT=10))

#Ejemplo optimizacion DEoptim
library(DEoptim)

DEctrl <- DEoptim.control(itermax = 500, trace=TRUE,
parallelType=1, packages=c(),
parVar=c("N", "bici", "metro", "micro", "auto", "choice", "TMAX", "alfa"))

mymleDE <- DEoptim(loglikel, lower=c(-10, -10, -10, -10, -10, 0.0001, 0.001),
upper=c(10, 10, 10, 10, 10, 10, 1), DEctrl)

```

# Bibliografía

- [1] György Abay. *Nachfrageabschätzung Swissmetro : eine Stated-Preference-Analyse, Nationales Forschungsprogramm 41, "Verkehr und Umwelt -. Wechselwirkungen Schweiz-Europa "*. Bern: EDMZ, 1999.
- [2] David Ardia, Kris Boudt, Peter Carl, K.M. Mullen y Brian Peterson. «Differential Evolution with DEoptim». En: *R Journal* 3 (ene. de 2011), págs. 27-34. DOI: 10 . 32614/RJ-2011-005.
- [3] Prawira F Belgiawan, Basil Schmid, Ilka Dubernet y Kay W Axhausen. «Comparison between RUM, RRM variants, and RAM. Swiss SP and RP data sets». en. En: 17th Swiss Transport Research Conference (STRC 2017); Conference Location: Ascona, Switzerland; Conference Date: May 17-19, 2017. Zurich: STRC, mayo de 2017. DOI: 10.3929/ethz-b-000130764.
- [4] Moshe E Ben-Akiva. «Structure of passenger travel demand models.» Tesis doct. Massachusetts Institute of Technology, 1973.
- [5] Moshe E Ben-Akiva y Steven R Lerman. «Some estimation results of a simultaneous model of auto ownership and mode choice to work». En: *Transportation* 3.4 (1974), págs. 357-376.
- [6] Moshe E Ben-Akiva, Steven R Lerman y Steven R Lerman. *Discrete choice analysis: theory and application to travel demand*. Vol. 9. MIT press, 1985.
- [7] Moshe Ben-Akiva, André de Palma, Daniel McFadden, Maya Abou-Zeid, Pierre-André Chiappori, Matthieu de Lapparent, Steven N. Durlauf, Mogens Fosgerau, Daisuke Fukuda, Stephane Hess, Charles Manski, Ariel Pakes, Nathalie Picard y Joan Walker. «Process and context in choice models». En: *Marketing Letters* 23.2 (jun. de 2012), págs. 439-456. DOI: 10.1007/s11002-012-9180-7.
- [8] Hasida Ben Zur y Shlomo J Breznitz. «The effect of time pressure on risky choice behavior». En: *Acta Psychologica* 47.2 (1981), págs. 89 -104. ISSN: 0001-6918. DOI: [https://doi.org/10.1016/0001-6918\(81\)90001-9](https://doi.org/10.1016/0001-6918(81)90001-9).

- [9] Christoph Bergmeir, Daniel Molina y José Benítez. «Memetic Algorithms with Local Search Chains in R: The Rmallschains Package». En: *Journal of Statistical Software, Articles* 75.4 (2016), págs. 1-33. ISSN: 1548-7660. DOI: 10.18637/jss.v075.i04.
- [10] Nicolas A J Berkowitsch, Benjamin Scheibehenne, Jörg Rieskamp y Max Matthäus. «A generalized distance function for preferential choices». En: *British Journal of Mathematical and Statistical Psychology* 68.2 (2015), págs. 310-325. DOI: 10.1111/bmsp.12048.
- [11] Nicolas A J Berkowitsch, Benjamin Scheibehenne y Jörg Rieskamp. «Rigorously testing multialternative decision field theory against random utility models.» En: *Journal of Experimental Psychology: General* 143.3 (2014), págs. 1331-1348. DOI: 10.1037/a0035159.
- [12] Michel Bierlaire. *PandasBiogeme: a short introduction*. Inf. téc. TRANSP-OR 181219. Transport y Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, 2018.
- [13] Michel Bierlaire, Kay Axhausen y Georg Abay. «The acceptance of modal innovation: The case of Swissmetro». En: (ene. de 2001).
- [14] Jerome R. Busemeyer. «Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models.» En: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11.3 (1985), págs. 538-564. DOI: 10.1037/0278-7393.11.3.538.
- [15] Jerome R Busemeyer y James T Townsend. «Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment.» En: *Psychological Review* 100.3 (jul. de 1993), págs. 432-459. DOI: 10.1037/0033-295X.100.3.432.
- [16] Colin F. Camerer. «An experimental test of several generalized utility theories». En: *Journal of Risk and Uncertainty* 2.1 (abr. de 1989), págs. 61-104. ISSN: 1573-0476. DOI: 10.1007/BF00055711.
- [17] Caspar G Chorus. «A New Model of Random Regret Minimization». En: *European Journal of Transport and Infrastructure Research* 10.2 (2010). ISSN: 1567-7141. DOI: 10.18757/ejtir.2010.10.2.2881.
- [18] Caspar G. Chorus y Michel Bierlaire. «An empirical comparison of travel choice models that capture preferences for compromise alternatives». En: *Transportation* 40.3 (mayo de 2013), págs. 549-562. ISSN: 1572-9435. DOI: 10.1007/s11116-012-9444-3.
- [19] Caspar G Chorus, Theo A Arentze y Harry J P Timmermans. «A Random Regret-Minimization model of travel choice». En: *Transportation Research Part B: Methodological* 42.1 (2008), págs. 1 -18. ISSN: 0191-2615. DOI: <https://doi.org/10.1016/j.trb.2007.05.004>.



- [20] J F Dashiell. *Fundamentals of general psychology*. Oxford, England: Houghton, Mifflin, 1937, págs. xxiv, 656-xxiv, 656.
- [21] Adele Diederich. «Dynamic Stochastic Models for Decision Making under Time Constraints». En: *Journal of Mathematical Psychology* 41.3 (1997), págs. 260 -274. ISSN: 0022-2496. DOI: <https://doi.org/10.1006/jmps.1997.1167>.
- [22] Ward Edwards. «The prediction of decisions among bets.» En: *Journal of Experimental Psychology* 50.3 (1955), págs. 201-214. ISSN: 0022-1015(Print). DOI: 10.1037/h0041692.
- [23] J F Engel, D T Kollat y R D Blackwell. «Consumer Behaviour». En: *New York: Holt, Rinehart, and Winston*. (Nov. de 1968).
- [24] C Angelo Guevara, Yue Tang y Song Gao. «The initial condition problem with complete history dependency in learning models for travel choices». En: *Transportation Research Part B: Methodological* 117 (nov. de 2018), págs. 850-861. DOI: 10.1016/j.trb.2017.09.006.
- [25] Thomas O Hancock, Stephane Hess y Charisma F Choudhury. «Decision field theory: Improvements to current methodology and comparisons with standard choice modelling techniques». En: *Transportation Research Part B: Methodological* 107 (ene. de 2018), págs. 18-40. DOI: <https://doi.org/10.1016/j.trb.2017.09.006>.
- [26] Maria Pilar Henriquez. «Desarrollo de modelo lexicográfico de ordenamiento parcial de atributos». Tesis de maestría. Departamento de Ingeniería Industrial. Universidad de Chile, 2014.
- [27] Stephane Hess. *From brainwaves to utilities: what can choice modellers learn from neuroscience and mathematical psychology?* Presentation at Interdisciplinary Choice Workshop (ICW), Santiago, Chile. Ago. de 2018.
- [28] Stephane Hess y David Palma. «Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application». En: *Journal of Choice Modelling* 32 (2019), pág. 100170. ISSN: 1755-5345. DOI: <https://doi.org/10.1016/j.jocm.2019.100170>.
- [29] Stephane Hess y David Palma. *Apollo user manual*. English. Ver. 0.0.8. 2019.
- [30] John D Hey, Gianna Lotito y Anna Maffioletti. «The descriptive and predictive adequacy of theories of decision making under uncertainty/ambiguity». En: *Journal of Risk and Uncertainty* 41.2 (oct. de 2010), págs. 81-111. ISSN: 1573-0476. DOI: 10.1007/s11166-010-9102-0.

- [31] Jared Hotelling, Jerome Busemeyer y Jiyun Li. «Theoretical Developments in Decision Field Theory: Comment on Tsetsos, Usher, and Chater (2010)». En: *Psychological review* 117 (oct. de 2010), págs. 1294-8. DOI: 10.1037/a0020401.
- [32] John S Hurland y Don N Kleinmuntz. «Factors influencing the use of internal summary evaluations versus external information in choice». En: *Journal of Behavioral Decision Making* 7.2 (1994), págs. 79-102. DOI: 10.1002/bdm.3960070202.
- [33] Donald Jamieson y William Petrusic. «Preference and time to choose». En: *Organizational Behavior and Human Performance* 19 (jun. de 1977), págs. 56-67. DOI: 10.1016/0030-5073(77)90054-X.
- [34] Kamel Jedidi, Rajeev Kohli y Ricardo Montoya. «Probabilistic lexicographic choice (No publicado)». En: *Columbia University* (2015).
- [35] Ran Kivetz, Oded Netzer y V Srinivasan. «Alternative models for capturing the compromise effect». En: *Journal of marketing research* 41.3 (2004), págs. 237-257.
- [36] Waiyan Leong y David A Hensher. «Contrasts of relative advantage maximisation with random utility maximisation and regret minimisation». En: *Journal of Transport Economics and Policy (JTEP)* 49.1 (2015), págs. 167-186.
- [37] Sarah Lichtenstein y Paul Slovic. «Response-induced reversals of preference in gambling: An extended replication in Las Vegas.» En: *Journal of Experimental Psychology* 101.1 (1973), págs. 16-20. ISSN: 0022-1015(Print). DOI: 10.1037/h0035472.
- [38] John Lynch y Thomas K. Srull. «Memory and Attentional Factors in Consumer Choice: Concepts and Research Methods». En: *Journal of Consumer Research* 9 (jun. de 1982), págs. 18-37. DOI: 10.1086/208893.
- [39] AA Markov. «Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain, The Notes of the Imperial Academy of Sciences of St». En: *Petersburg VIII Series, Physio-Mathematical College* 22.9 (1907).
- [40] Daniel McFadden. «Rationality for Economists?» En: *Journal of Risk and Uncertainty* 19.1 (dic. de 1999), págs. 73-105. ISSN: 1573-0476. DOI: 10.1023/A:1007863007855.
- [41] Daniel McFadden. «The Revealed Preferences of a Government Bureaucracy: Empirical Evidence». En: *The Bell Journal of Economics* 7.1 (1976), págs. 55-72. ISSN: 0361915X.
- [42] Daniel McFadden y Kenneth Train. «Mixed MNL models for discrete response». En: *Journal of Applied Econometrics* 15.5 (2000), págs. 447-470. DOI: 10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/1099-1255%28200009/10%2915%3A5%3C447%3A%3AAID-JAE570%3E3.0.CO%3B2-1>.

- [43] D. Molina, M. Lozano, C. García-Martínez y F. Herrera. «Memetic Algorithms for Continuous Optimisation Based on Local Search Chains». En: *Evolutionary Computation* 18.1 (mar. de 2010), págs. 27-63. ISSN: 1063-6560. DOI: 10.1162/evco.2010.18.1.18102.
- [44] Henry Montgomery y Ola Svenson. «On Decision Rules and Information Processing Strategies for Choices Among Multiattribute Alternatives». En: *Scandinavian Journal of Psychology* 17 (mayo de 1976), págs. 283 -291. DOI: 10.1111/j.1467-9450.1976.tb00241.x.
- [45] Frederick Mosteller y Philip Nogee. «An Experimental Measurement of Utility». En: *Journal of Political Economy* 59 (1951).
- [46] Katharine Mullen. «Continuous Global Optimization in R». En: *Journal of Statistical Software, Articles* 60.6 (2014), págs. 1-45. ISSN: 1548-7660. DOI: 10.18637/jss.v060.i06.
- [47] Takao Noguchi y Neil Stewart. «In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions». En: *Cognition* 132.1 (2014), págs. 44 -56. ISSN: 0010-0277. DOI: <https://doi.org/10.1016/j.cognition.2014.03.006>.
- [48] John W Payne. «Heuristic Search Processes in Decision Making». En: *Advances in Consumer Research* 03.1 (1976), págs. 321-327.
- [49] John W Payne, James R Bettman y Eric J Johnson. «Behavioral Decision Research: A Constructive Processing Perspective». En: *Annual Review of Psychology* 43.1 (1992), págs. 87-131. DOI: 10.1146/annurev.ps.43.020192.000511.
- [50] John W Payne, James R Bettman y Eric J Johnson. *The Adaptive Decision Maker*. Cambridge University Press, 1993. DOI: 10.1017/CB09781139173933.
- [51] Kenneth Price, Rainer M. Storn y Jouni A. Lampinen. *Differential Evolution A Practical Approach to Global Optimization*. Springer-Verlag, 2006. DOI: 10.1007/3-540-31306-0.
- [52] Huanmei Qin, Hongzhi Guan y Yao-Jan Wu. «Analysis of park-and-ride decision behavior based on Decision Field Theory». En: *Transportation Research Part F: Traffic Psychology and Behaviour* 18 (2013), págs. 199 -212. ISSN: 1369-8478. DOI: 10.1016/j.trf.2013.02.001.
- [53] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2017.
- [54] RStudio Team. *RStudio: Integrated Development Environment for R*. Ver. V 1.2.1335. RStudio, Inc. Boston, MA, 2015.

- [55] Robert Roe, Jerome Busemeyer y James Townsend. «Multialternative decision field theory: A dynamic connectionst model of decision making». En: *Psychological review* 108 (abr. de 2001), págs. 370-92. DOI: 10.1037/0033-295X.108.2.370.
- [56] Barry Schwartz, Andrew Ward, John Monterosso, Sonja Lyubomirsky, Katherine White y Darrin R Lehman. «Maximizing versus satisficing: Happiness is a matter of choice.» En: *Journal of Personality and Social Psychology* 83.5 (2002), págs. 1178-1197. DOI: 10.1037/0022-3514.83.5.1178.
- [57] Savannah Wei Shi, Michel Wedel y F. G. M. (Rik) Pieters. «Information Acquisition During Online Decision Making: A Model-Based Exploration Using Eye-Tracking Data». En: *Management Science* 59.5 (2013), págs. 1009-1026. DOI: 10.1287/mnsc.1120.1625. eprint: <https://doi.org/10.1287/mnsc.1120.1625>.
- [58] Herbert A Simon. *Models of man; social and rational*. Oxford, England: Wiley, 1957, págs. xiv, 287-xiv, 287.
- [59] Paul Slovic. «Choice between equally valued alternatives.» En: *Journal of Experimental Psychology: Human Perception and Performance* 1.3 (1975), págs. 280-287. DOI: 10.1037/0096-1523.1.3.280.
- [60] Peter Stüttgen, Peter Boatwright y Robert T. Monroe. «A Satisficing Choice Model». En: *Marketing Science* 31.6 (2012), págs. 878-899. DOI: 10.1287/mksc.1120.0732. eprint: <https://doi.org/10.1287/mksc.1120.0732>.
- [61] L L Thurstone. «A law of comparative judgment.» En: *Psychological Review* 34.4 (1927), págs. 273-286. DOI: 10.1037/h0070288.
- [62] Amos Tversky. «Elimination by aspects: A theory of choice.» En: *Psychological Review* 79.4 (1972), págs. 281-299. DOI: 10.1037/h0032955.
- [63] Amos Tversky e Itamar Simonson. «Context-dependent preferences». En: *Management science* 39.10 (1993), págs. 1179-1189.
- [64] Thomas S. Wallsten y Curtis Barton. «Processing probabilistic multidimensional information for decisions.» En: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8.5 (1982), págs. 361-384. DOI: 10.1037/0278-7393.8.5.361.
- [65] Yifei Xie. «Choice Modeling with Observed and Unobserved Information Search». Tesis de maestría. Department of Civil Engineering. Massachusetts Institute of Technology, jun. de 2019.
- [66] Liu (Cathy) Yang, Olivier Toubia y Martijn G. De Jong. «A Bounded Rationality Model of Information Search and Choice in Preference Measurement». En: *Journal of Marketing Research* 52.2 (2015), págs. 166-183. DOI: 10.1509/jmr.13.0288. eprint: <https://doi.org/10.1509/jmr.13.0288>.