

Modelo de maximización de la entropía y costos generalizados intervalares para la distribución de viajes urbanos¹

**Maximization Model of entropy and generalized costs
for urban travel distribution²**

**Modelo de maximização da entropia e custos generalizados
intervalares para a distribuição de viagens urbanas³**

Héctor López-Ospina⁴

SICI: 0123-2126(201307)17:2<391:MMECGI>2.0.TX;2-L

¹ Fecha de recepción: 15 de febrero de 2012. Fecha de aceptación: 15 de septiembre de 2013. Universidad de Chile, Santiago de Chile, Chile.

² Received: February 15, 2012. Accepted: September 15, 2013. Universidad de Chile, Santiago de Chile, Chile.

³ Data de recebimento: 15 de fevereiro de 2012. Data de aceitação: 15 de setembro de 2013. Universidade de Chile, Santiago de Chile, Chile.

⁴ Matemático y magíster en Matemática Aplicada, de la Universidad Nacional de Colombia. PhD. en Sistemas de Ingeniería, Universidad de Chile. Especialidad en Transporte, Universidad de Chile, Santiago de Chile, Chile. Correo electrónico: hlopezospina@ing.uchile.cl.

Resumen

En este trabajo se propone la formulación de un modelo de distribución de viajes urbanos doblemente acotado en origen y destino, basado en la maximización de la entropía y minimización de los costos generalizados de viajes, relajando la condición de que dichos costos son un parámetro, extendiendo la formulación bajo el supuesto de que son variables difusas con función de pertenencia intervalar. Adicionalmente, dados dichos supuestos, se obtiene una formulación de un modelo de optimización asociado con elecciones discretas, y se presenta una propuesta de uso del modelo en predicciones. Se desarrolla un ejemplo numérico para explicar la teoría desarrollada.

Palabras clave

Distribución de viajes, maximización de la entropía, optimización difusa, costos intervalares.

Abstract

In this paper we propose the formulation of a model for urban double travel distribution delimited at origin and destination, based on the maximization of entropy and minimization of the generalized travel costs by relaxing the condition that such costs are a parameter, extending the formulation under the assumption that they are fuzzy variables with an intervalar function. In addition, given such assumptions, we propose a formulation of an optimization model associated with discrete choices, and we present a proposal for using the model in predictions. A numerical example is developed for explaining the theory developed.

Keywords

Travel distribution, maximization of entropy, fuzzy optimization, intervalar costs.

Resumo

Neste trabalho propõe-se formulação de um modelo de distribuição de viagens urbanas limitado na dupla origem destino, baseado na maximização da entropia e minimização dos custos generalizados de viagens, relaxando a condição de tais custos serem um parâmetro, estendendo a formulação sob a suposição de que são variáveis difusas com função de pertinencia intervalar. Adicionalmente, dados tais supostos, obtém-se formulação de um modelo de otimização associado com eleições discretas, e apresenta-se uma proposta de uso do modelo em predições. Apresenta-se exemplo numérico para explicar a teoria desenvolvida.

Palavras-chave

Distribuição de viagens, maximização da entropia, otimização difusa, custos intervalares.

1. Introducción

En muchos problemas prácticos asociados con la solución de un modelo de optimización matemática como soporte para la toma de decisiones, no se puede considerar el conjunto de parámetros o datos como completamente conocido o determinado, debido a que en diversas aplicaciones dichos datos varían en forma significativa por medio de la experimentación y análisis, o simplemente de acuerdo con la experiencia del tomador de decisiones. Un ejemplo de esto son los costos generalizados de transporte asociados con un viaje entre dos zonas de un sistema urbano, que incluyen los costos asociados con la valoración del tiempo de espera, de acceso y de viaje de cada modo, con la inclusión de la tarifa entre dos zonas de una ciudad (Ortúzar y Willumsen, 2008), que pueden variar de acuerdo con el tipo de viaje, con la persona que realiza el viaje o el horario. Es más, pueden existir variaciones temporales de acuerdo con los cambios en factores exógenos a los individuos, como la frecuencia de los buses, la congestión, el clima, etc.

Para encontrar solución a este tipo de problemas y generar resultados más robustos en diversas áreas de ingeniería y ciencias, se han utilizado varios métodos y técnicas matemáticas y estadísticas. Dentro de los más conocidos se encuentran la optimización estocástica o probabilística, donde se supone que el conjunto de parámetros son aleatorios, pero se tienen distribuciones de probabilidad asociadas. Por otro lado, se tiene la optimización difusa, donde la imprecisión de los datos se puede representar por medio de un grado de pertenencia difusa a un conjunto, ya que los objetos de estudio pueden pertenecer a varias clases (Cadenas y Verdegay, 1999).

Este trabajo se basa en la teoría de los modelos de distribución de viajes urbanos o de un sistema interconectado, donde se supondrá que los costos generalizados tienen como parámetro un intervalo de pertenencia asociado con las posibles variaciones de estos valores, lo que permite generar un análisis de sensibilidad a priori de la solución del modelo e incluye robustez en las solucio-

nes factibles. De esta forma, se propone resolver por medio de una formulación como un problema de optimización de elecciones discretas, para obtener conclusiones válidas sobre las variaciones de los parámetros. La sección 2 explica brevemente el problema de máxima entropía aplicado a la teoría de la probabilidad y desde el punto de vista de la combinatoria. En la sección 3 se desarrolla la formulación del problema y su solución como un problema de elecciones discretas, y por último, en la sección 4, se desarrolla un ejemplo para describir el modelo desarrollado.

2. Maximización de la entropía

Basados en Fang *et al.* (1997) y Romero (2008) se presentan los principios básicos de la máxima entropía, desde el punto de vista probabilístico, asumiendo distribuciones discretas y finitas. El objetivo de la máxima entropía es generar distribuciones de probabilidad con información dada. De forma más exacta, entre todas las distribuciones de probabilidad que son factibles en un conjunto finito de estados bajo cierto tipo de información, se desea encontrar la que soporte mayor nivel de incertidumbre. Matemáticamente, se define entropía de la siguiente forma:

Definición: Sea X una variable aleatoria discreta, con valores de los estados x_1, x_2, \dots, x_n y con distribución $p_i = P(X = x_i)$. Se define la entropía de Shannon para la distribución $p = (p_1, \dots, p_n)$ como $H(p) := -\sum_i p_i \ln(p_i)$, (1).

La función $H(p)$ evalúa la cantidad de información que se espera obtener al conocer el valor de X . A su vez, mide el nivel de incertidumbre asociado con próximas realizaciones de X . Como ejemplo, tomemos una variable aleatoria Bernoulli con parámetro p , entonces la entropía se define como: $H(p) := -p \ln(p) - (1-p) \ln(1-p)$. En el caso en que p tiende a un valor extremo 0 o 1, entonces la solución tenderá a 0, indicando menor estocasticidad en el proceso. Por otro lado, la función de entropía alcanza su valor máximo en $p = 0,5$, donde se tiene el mayor nivel de incertidumbre posible; es decir, cuando las dos opciones tienen igual probabilidad de ocurrencia. Shannon demuestra que la función $H(p)$ es la única que satisface una cantidad de propiedades que definen la entropía en una distribución (Shannon, 1948). Claramente, en cualquier problema de maximización de la entropía, la solución óptima es la distribución uniforme. Es decir $p_1 = p_2 = \dots = p_n = \frac{1}{n}$, pero en el caso de tener información adicional del sistema, dicha distribución $\frac{1}{n}$ puede variar de acuerdo con los datos a priori obtenidos. Matemáticamente, si se supone que se tiene $g_2(X), \dots, g_m(X)$ m funciones de X

con valor esperado conocido notado como a_1, a_2, \dots, a_m , entonces la distribución determinada por el problema de máxima entropía con la información asociada con las funciones $g_j(X)$ es:

$$\max H(p): = -\sum_i p_j \ln(p_j) \text{ Sujeto a: } \sum_{j=1}^n p_j g_j(x_j) = a_i, i = 1, \dots, m \sum_{j=1}^n p_j = 1, p_j \geq 0 \tag{2}$$

Y a su vez, la solución del problema de optimización (ecuación 2) determina una distribución consistente con las restricciones y que maximiza la incertidumbre. Este tipo de formulaciones se puede extender a restricciones no lineales, como cuadráticas o de entropía (Fang *et al.*, 1997), pero dado el objetivo del presente trabajo, solo se asumirán restricciones de tipo lineal. Por otro lado, cuando se conocen distribuciones a priori notadas como $p^0 = (p_1^0, \dots, p_n^0)$, es factible variar la formulación de la función objetivo buscando maximizar $-\sum_i p_j \ln(p_j / p_j^0)$.

Existe otro argumento para la derivación de la entropía, usando modelos combinatorios que se asocian con el cálculo de la entropía de una distribución y está formalmente descrito en Romero (2008), y en particular en Ortúzar *et al.* (2008), aplicado al problema de distribución de viajes, donde se demuestra que la distribución más probable de viajes entre dos zonas i, j $T_{ij} = T * p_{ij}$ donde T es el total de viajes generados o atraídos y p_{ij} es la probabilidad de viajar, entre el par origen destino (i, j) está dado por la maximización de la ecuación de Wilson (1970), Ortúzar *et al.* (2008):

$$W(T_{ij}) = \frac{T!}{\prod_{(i,j)} T_{ij}!} \tag{3}$$

Por conveniencia, aplicando una transformación monótona creciente a la ecuación 3, por ejemplo el logaritmo natural, se obtiene la misma solución óptima; entonces, la distribución más probable es la solución del siguiente problema: $\ln W(T_{ij}) = \ln T! - \sum_{(i,j)} \ln(T_{ij}!)$. Usando la aproximación de Stirling $\log X! \approx X * \log X - X$, (Romik, 2000), sustituyendo y eliminando los términos que no son optimizables, se obtiene el siguiente problema de programación matemática:

$$\text{Max}_{T_{ij}} - \sum_{i,j} T_{ij} \ln T_{ij} \text{ Sujeto a } \sum_{i,j} T_{ij} = T \tag{4}$$

El problema anterior tiene como solución la distribución uniforme de viajes, debido a que esta es la que genera mayor nivel de incertidumbre, pero en el caso de conocer algún tipo de información del sistema, entonces la solución podría

ser distinta, por ejemplo, información acerca de los viajes originados o atraídos por zona, o costos, o tiempo de viaje, obteniendo resultados de distribuciones como los mostrados para el problema de probabilidad. Este caso en particular se explicará en la siguiente sección.

3. Planteamiento del problema

En esta sección se realizará una breve introducción al problema de máxima entropía aplicado a la distribución de viajes, suponiendo costos deterministas y extensión a costos intervalares. Esta primera parte está basada en Ortúzar y Willumsen (2008), y Fang *et al.* (1997). La planeación en los sistemas de transporte urbano se considera un proceso de cuatro etapas o fases, definidas así:

- Etapa 1: generación y atracción de viajes.
- Etapa 2: distribución de viajes.
- Etapa 3: partición modal.
- Etapa 4: asignación en la red.

En estos procesos se considera que la ciudad está dividida en zonas disyuntas, y uno de los objetivos importantes es estimar el número de viajes entre zonas en un periodo determinado (por ejemplo, 24 horas), lo cual es conocido como el sistema de distribución de viajes (etapa 2). Por lo general, el número de viajes que se generan en una zona o el número de viajes que atrae una zona son estimados por modelos de regresión lineal múltiple, por medio de encuestas basadas en las características de los hogares (número de personas, número de trabajadores, ingreso, número de estudiantes, etc.) o el uso del suelo de la zona (cantidad de comercio, centros educativos, empresas, etc.). Estos resultados de generación y atracción por zona son necesarios para estimar su distribución. Uno de los modelos clásicos usados para obtener la distribución de viajes T_{ij} , donde i es la zona origen y j es la zona de destino, es el de máxima entropía doblemente acotada, que matemáticamente se describe como en la ecuación 5 (Ortúzar y Willumsen, 2008):

$$\begin{aligned}
 & \text{Max}_{T_{ij}} - \sum_{i=1}^K \sum_{j=1}^L T_{ij} \ln T_{ij} \\
 & \text{Sujeto a:} \\
 & \sum_{j=1}^L T_{ij} = O_i, \quad i = 1, \dots, K; \\
 & \sum_{i=1}^K T_{ij} = D_j, \quad j = 1, \dots, L; \\
 & \sum_{i=1}^K \sum_{j=1}^L c_{ij} T_{ij} = C; \quad T_{ij} \geq 0
 \end{aligned} \tag{5}$$

Donde la función objetivo del problema (ecuación 5) define la entropía en el sistema o el grado de incertidumbre del sistema urbano asociado, O_i son los viajes generados por la zona i , D_j es la cantidad de viajes atraídos por la zona j , c_{ij} es el costo generalizado de un viaje por unidad entre el par origen destino (i,j) y C es el costo total del sistema de transporte que asumen todos los usuarios. Tanto c_{ij} y C se explican de forma más detallada en la siguiente subsección. Además, si se transforma cada variable como $p_{ij} = T_{ij}/T$, donde T es el total de viajes generados o atraídos en el sistema, entonces se obtiene una formulación equivalente a la descrita en la sección anterior, donde p_{ij} es la probabilidad de realizar un viaje entre la zona i y la zona j .

3.1. Costos generalizados

En esta sección se describe el concepto de costo generalizado de viaje entre un par origen destino en un sistema de transporte urbano y la razón para analizar dicho parámetro como un intervalo. Basándose en Ortúzar y Willumsen (2008), se pueden definir los costos generalizados en un sistema de transporte como una medida o valor que representa la combinación de los atributos más importantes o relevantes asociados con la desutilidad de realizar un viaje entre un par origen destino (i,j) . El costo generalizado de viaje entre un par origen destino (i,j) se puede representar como una función lineal de los atributos propios del viaje ponderados con valores que representan la importancia o valoración de cada uno de estos atributos o características, y que son percibidos por el agente que realiza el viaje. Esas valoraciones se obtienen a través de la teoría de elecciones discretas y se pueden suponer difusas o intervalares por los cambios en la valoración de los agentes o variaciones del sistema, tal como se hace en Momeni y Nazemi (2010) para un problema de elección de vehículo o en Dell’Orco y Kikuchi (2004), que, basados en la teoría de la posibilidad para la comparación de utilidades, plantean un modelo de elección de ruta.

Se puede suponer que el costo generalizado intervalar asociado con un modo de viaje k en el par origen destino (i,j) se escribe como en la ecuación 6:

$$C_{ij}^k = A_1 \otimes T_{ij}^v \oplus A_2 \otimes T_{ij}^c \oplus A_3 \otimes T_{ij}^e \oplus A_4 \otimes T_{nij} \oplus A_5 \otimes F_{ij} \oplus A_6 \otimes \phi_j + \delta \quad (6)$$

- Cada componente de dichos costos se explica a continuación:
- $\otimes \oplus$ son las notaciones usadas para el producto y suma de intervalos cerrados y acotados.

- $T_{ij}^v = [t_{ij}^{v,\min}, t_{ij}^{v,\max}]$ es el intervalo asociado con el tiempo de viaje factible a bordo del vehículo para ir de i a j , donde $t_{ij}^{v,\min}$ es el tiempo mínimo y $t_{ij}^{v,\max}$ es el tiempo máximo.
- $T_{ij}^c = [t_{ij}^{c,\min}, t_{ij}^{c,\max}]$ es el intervalo asociado con el tiempo de caminata posible para acceder al modo y llegar al destino.
- $T_{ij}^e = [t_{ij}^{e,\min}, t_{ij}^{e,\max}]$ es el intervalo asociado con el tiempo de espera viable.
- $T_{nij} = [t_{nij}^{\min}, t_{nij}^{\max}]$ es el intervalo asociado con el tiempo de transbordo.
- ϕ_j es un costo terminal que se supondrá determinista, asociado con estacionamiento en el destino.
- $F_{ij} = [f_{ij}^{\min}, f_{ij}^{\max}]$ es el intervalo que recoge el costo del viaje.
- $A_1, A_2, A_3, \dots, A_6$ son los intervalos donde se encuentran las valoraciones asociadas con cada uno de los elementos del viaje.

Vale la pena anotar que en este caso, el intervalo C_{ii}^k asociado con un costo generalizado de viaje es obtenido por medio de la aritmética intervalar. De esta forma, es importante tener en cuenta las siguientes propiedades de la lógica difusa.

Definición: Dados dos intervalos cerrados y acotados $X = [X_1, X_2]$ y $Y = [Y_1, Y_2]$, entonces la suma y el producto entre X y Y se definen como: $X \oplus Y = [X_1 + Y_1, X_2 + Y_2]$, y si se toman X_1, X_2, Y_1, Y_2 positivos no nulos, $X \otimes Y = [X_1 * Y_1, X_2 * Y_2]$.

Determinar el costo generalizado en términos intervalares permite analizar la sensibilidad de variaciones en los atributos y su implicancia en los resultados en los modelos de distribución de viajes. Además, no es necesario tener un consolidado de los datos asociados con los costos y tiempo de viaje, sino estimaciones de las valoraciones extremas.

3.2 Solución del problema como un problema de elecciones discretas

Teniendo en cuenta el supuesto de que el costo generalizado de viajar entre el par origen destino es un intervalo $C_{ij} = [C_{ij}^{\min}, C_{ij}^{\max}]$, entonces el costo total del sistema C también varía en un rango de posibilidades que se notará en forma de intervalo como $C = [C^{\min}, C^{\max}]$. De esta forma, el problema de optimización con entropía con costos intervalares se escribe de la forma:

$$\max_{T_{ij}} - \sum_{i=1}^K \sum_{j=1}^L T_{ij} \ln T_{ij}$$

Sujeto a:

$$\begin{aligned} \sum_{j=1}^L T_{ij} &= O_i, i = 1, \dots, K: \\ \sum_{i=1}^K T_{ij} &= D_j, j = 1, \dots, L \\ \sum_{i=1}^K \sum_{j=1}^L [C_{ij}^{\min}, C_{ij}^{\max}] * T_{ij} &= [C^{\min}, C^{\max}] \end{aligned} \tag{7}$$

Teniendo en cuenta que cualquier $C_{ij}^l \in [C_{ij}^{\min}, C_{ij}^{\max}]$ se puede escribir como una combinación convexa de C_{ij}^{\min} y de la forma $C_{ij}^l = \lambda C_{ij}^{\max} + (1 - \lambda) C_{ij}^{\min}$ con $\lambda \in [0, 1]$, entonces la ecuación 7 se puede representar como:

$$\begin{aligned} \max_{T_{ij}} & - \sum_{i=1}^K \sum_{j=1}^L T_{ij} \ln T_{ij}, \\ \text{Sujeto a:} \\ \sum_{j=1}^L T_{ij} &= O_i, i = 1, \dots, K \\ \sum_{i=1}^K T_{ij} &= D_j, j = 1, \dots, L \\ \sum_{i=1}^K \sum_{j=1}^L (\lambda C_{ij}^{\max} + (1 - \lambda) C_{ij}^{\min}) * T_{ij} &= \lambda C^{\max} + (1 - \lambda) C^{\min}, \lambda \in [0, 1] \end{aligned} \tag{8}$$

Para resolver el problema anterior se usará una aproximación basada en la discretización del valor de λ , dado que asumirla como una variable continua implicaría una complejidad bastante alta en la solución, debido a la no linealidad que se generaría en la restricción de costos. Es decir se resolverán $n + 1$ problemas de entropía asumiendo los siguientes valores para $\lambda = \left\{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\right\}$, donde $\lambda = 0$ implica que los agentes asumen costos mínimos, $\lambda = 1$ indica la consideración que se asumen costos máximos y $\frac{k}{n}$ indica las opciones intermedias. De esta forma, se notará el problema k -ésimo de entropía para la distribución de viajes como:

$$\begin{aligned} \max_{T_{ij}} & - \sum_{i=1}^K \sum_{j=1}^L T_{ij} \ln T_{ij}, \\ \text{Sujeto a:} \\ \sum_{j=1}^L T_{ij} &= O_i, i = 1, \dots, K: \\ \sum_{i=1}^K T_{ij} &= D_j, j = 1, \dots, L \\ \sum_{i=1}^K \sum_{j=1}^L (\lambda_k C_{ij}^{\max} + (1 - \lambda_k) C_{ij}^{\min}) * T_{ij} &= \lambda_k C^{\max} + (1 - \lambda_k) C^{\min} \end{aligned} \tag{9}$$

Donde $\lambda_k = \frac{k}{n}, k = 0, 1, \dots, n$. Además, se notará $C_k = \lambda_k C^{\max} + (1 - \lambda_k) C^{\min}$.

Por otro lado, se propone el uso de valores intermedios de los costos generalizados como escenarios posibles de optimización, ya que la no linealidad de la

función de entropía no garantiza que la solución se encuentre en los extremos del intervalo.

Si se notan $\alpha_i, \gamma_j, \beta_k$ a los multiplicadores de Lagrange asociados con la ecuación 9, entonces se obtiene por condiciones de primer orden que:

$$T_{ijk} = \exp(\alpha_i) \exp(\gamma_j) \exp(-\beta_k (\lambda_k C_{ij}^{\max} + (1 - \lambda_k) C_{ij}^{\min}))$$

Remplazando $\exp(\alpha_i) = O_i A_i^k$ y $\exp(\gamma_j) = D_j B_j^k$ entonces

$$T_{ijk} = O_i A_i^k D_j B_j^k \exp(-\beta_k (\lambda_k C_{ij}^{\max} + (1 - \lambda_k) C_{ij}^{\min}))$$

Los valores de A_i^k y B_j^k miden el nivel de beneficios asociado con viajar desde i o a recibir un viaje en j , respectivamente asumiendo costos de viaje $(\lambda_k C_{ij}^{\max} + (1 - \lambda_k) C_{ij}^{\min})$. Por lo cual son medidas clave para calcular índices de accesibilidad y atractividad (Martínez, 1995; Martínez y Araya, 2000) en proyectos de transporte y uso de suelo. Por otro lado, β_k es la valoración del costo de viaje de los individuos. Nótese que la variación de λ_k implica la variación de A_i^k, B_j^k y β_k , lo que indica que es factible obtener un intervalo para cada medida. Este resultado es importante en la valoración de proyectos de transporte, pues genera análisis de sensibilidad o rangos factibles de dichas medidas.

La ecuación 9 se puede resolver por medio del sistema de puntos fijos generados por el sistema de ecuaciones, de acuerdo con las condiciones de primer orden. Las técnicas de calibración y estimación de los parámetros del problema de entropía se pueden revisar en Ortuzar y Willumsen (2008). Es importante tener en cuenta que para obtener A_i^k, B_j^k y β_k es necesario conocer la siguiente información de un año base o información de la matriz a priori que se describe a continuación:

- T_{ij}^0 , distribución inicial de viajes en el año base.
- O_{ij}^0, D_j^0 viajes generados y atraídos por zona.
- $C_{ij}^0(\lambda_k) = \lambda_k C_{ij}^{\max} + (1 - \lambda_k) C_{ij}^{\min}$ costos generalizados máximos y mínimos de viaje y su combinación convexa.
- $C_k^0 = \lambda_k C^{\max} + (1 - \lambda_k) C^{\min} = \sum_{i=1}^k \sum_{j=1}^L (\lambda_k C_{ij}^{\max} + (1 - \lambda_k) C_{ij}^{\min}) * T_{ij}^0$

Dado que $\lambda_k \in [0, 1]$, entonces con la información previa y la solución de los $n + l$ problemas de optimización no lineal (ecuación 9) se obtiene un intervalo que asume la variación de la valoración del costo de viaje de los individuos, notada como $\beta \in [\beta^{\min}, \beta^{\max}]$. Este intervalo explica el rango factible de la sen-

sibilidad al costo de viaje. Así como se mencionaron intervalos para las medidas de accesibilidad y los valores óptimos de T_{ij}^* .

Sin embargo, para predicciones se debe usar un solo valor de β , por lo que se propone usar aquel que genere la máxima entropía. Esto es debido a que la información determinística en la ecuación 9 es la cantidad de viajes generados y atraídos por zona (O_i, D_j, \forall_i, j) . Pero en los casos aplicados en la práctica, el costo total es una aproximación estimada por la información de la matriz a priori y el uso de los costos generalizados medios (por simplicidad, $\lambda = 0,5$). Sin embargo, el promedio es una medida poco robusta, ya que en casos de presencia de datos atípicos dicho valor se puede ver sesgado o mal estimado. Por otro lado, la formulación intervalar añade un componente probabilístico interesante, donde se puede suponer que el costo de viaje es una variable aleatoria uniforme en el intervalo $[C_{ij}^{\min}, C_{ij}^{\max}]$, por lo cual se tiene la misma probabilidad de que un individuo tomado al azar asuma cualquier costo en el intervalo; de esta forma, se propone que la solución del problema de máxima entropía con costos intervalares se obtiene de la forma:

$$\max_{\lambda_k} \left(\max_{i,j} - \sum_{i=1}^K \sum_{j=1}^L T_{ij} \ln T_{ij} \right)$$

Sujeto a:

$$\sum_{j=1}^L T_{ij} = O_i, i = 1, \dots, K$$

$$\sum_{i=1}^K T_{ij} = D_j, j = 1, \dots, L \tag{10}$$

$$\sum_{i=1}^K \sum_{j=1}^L (\lambda_k C_{ij}^{\max} + (1 - \lambda_k) C_{ij}^{\min}) * T_{ij} = \lambda_k C^{\max} + (1 - \lambda_k) C^{\min}$$

La ecuación 10 se puede analizar como un problema en dos etapas; en la primera se resuelve un problema de optimización no lineal paramétrico en λ_k y en la segunda etapa se encuentra la máxima entropía, dada la variación de dicho valor.

La ecuación 10 podría mejorar las predicciones que se realizan en el sector de transporte en la práctica, donde generalmente se usan costos promedio, ignorando la variabilidad intrínseca de dicha medida o la poca robustez, dada por la presencia de datos atípicos o aglomeraciones por sectores. El hecho de usar la entropía como criterio de selección permite determinar la distribución más probable dada una información conocida como O_i y D_j y bajo el supuesto de que cada costo en el intervalo de pertenencia tiene la misma probabilidad de ocurrencia.

4. Ejemplo numérico

Para ejemplificar el modelo propuesto con costos intervalares se va a trabajar con una zonificación agregada y basados en datos simulados. En particular, no se describe un cálculo de los costos intervalares, tal como fue descrito en la ecuación 6, debido a que se busca mostrar la bondad del modelo de optimización. En la tabla 1 se presenta la matriz de costos.

Tabla 1. Costos generalizados intervalares para el ejemplo numérico

Zonas	1	2	3	4
1	[2,5, 3]	[11, 12]	[18,5, 19]	[22, 24]
2	[12, 13]	[3, 4]	[13,5, 14]	[19,5, 21]
3	[15, 17]	[13,5, 15]	[4,5, 5]	[7, 8]
4	[24, 25]	[18,5, 19]	[8,5, 10]	[4,5 ,5]

Fuente: presentación propia de los autores.

Adicionalmente, los vectores orígenes destinos están dados por la información expuesta en la tabla 2.

Tabla 2. Viajes originados y atraídos por zona

Zona	O _i	D _i
1	355	205
2	455	355
3	255	455
4	570	620

Fuente: presentación propia de los autores.

Basados en esta información, el problema de distribución de viajes (ecuación 10) de la maximización de la entropía y costos paramétricos en λ_k se puede escribir de la siguiente forma extensiva:

$$\max_{T_{ij}} - \sum_{i=1}^4 \sum_{j=1}^4 T_{ij} \ln T_{ij}$$

Restricciones de balance en el origen:

$$\sum_{j=1}^4 T_{1j} = 355, \sum_{j=1}^4 T_{2j} = 455, \sum_{j=1}^4 T_{3j} = 255, \sum_{j=1}^4 T_{4j} = 570$$

Restricciones de balance en el destino:

$$\sum_{i=1}^4 T_{i1} = 205, \sum_{i=1}^4 T_{i2} = 355, \sum_{i=1}^4 T_{i3} = 455, \sum_{i=1}^4 T_{i4} = 620$$

Restricción de costos:

$$\{2,5(1-\lambda_k) + 3\lambda_k\}T_{11} + \{11(1-\lambda_k) + 12\lambda_k\}T_{12} + \{18,5(1-\lambda_k) + 19(\lambda_k)\}T_{13} + \dots + \{4,5(1-\lambda_k) + 5(\lambda_k)\}T_{44} = \lambda_k C^{max} + (1-\lambda_k) C^{min}$$

Donde C^{min} y C^{max} se obtienen con la información de la matriz de distribución a priori (T_{ij}^0) de la forma:

$$C^{min} = \sum_{ij} C_{ij}^{min} T_{ij}^0 \text{ y } C^{max} = \sum_{ij} C_{ij}^{max} T_{ij}^0$$

Para mostrar que la máxima entropía se puede alcanzar en cualquier valor del intervalo asociado al costo generalizado, se usarán dos matrices a priori distintas. En el primer caso, la máxima entropía se alcanza en el extremo inferior del intervalo de los costos; en el segundo, la máxima entropía se alcanza cuando se asume el extremo superior de dicho intervalo.

Matriz a priori 1: en la matriz de referencia o a priori (T_{ij}^0) se tiene mayor cantidad de viajes extrazonales que intrazonales (tabla 3).

Tabla 3. Matriz a priori 1

Zonas	1	2	3	4
1	25	100	60	170
2	40	25	60	330
3	60	60	25	110
4	80	170	310	10

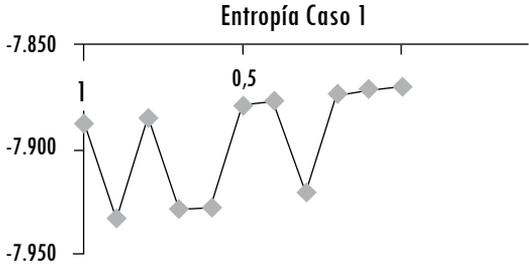
Fuente: presentación propia de los autores.

La figura 1 muestra la variación de la máxima entropía para diferentes valores de $\lambda_k \in \{0, 0,1, 0,2, \dots, 1\}$ usando los costos obtenidos por la matriz a priori 1. Los resultados fueron obtenidos usando una versión gratuita del *software* GAMS (López, 2006).

En este caso, la máxima entropía se alcanza en los costos mínimos ($\lambda = 0$), pero existen estructuras de costos intermedios (por ejemplo, $\lambda = 0,9$ o $0,7$) que generan menor entropía que el problema de valor extremo asociado con el máximo del intervalo ($\lambda = 1$), lo que implica que dada la información de la matriz

a priori, los viajes originados y atraídos por zona se tendría mayor probabilidad de obtener una distribución con la información asociada con los costos máximos que con algunas valoraciones intermedias.

Figura 1. Máxima entropía para diferentes valores de λ y la matriz a priori 1



Fuente: presentación propia de los autores.

Adicionalmente, la valoración del costo generalizado de viaje b_k varía dentro del siguiente intervalo $[-0,6787; -0,07208]$, donde β_k es el multiplicador de Lagrange de la solución del problema de entropía con λ_k

Matriz a priori 2: en este caso se realizan más viajes entre las mismas zonas, con lo cual se tiene la matriz a priori (T_{ij}^0) de la tabla 4.

Tabla 4. Matriz a priori 2

Zonas	1	2	3	4
1	200	20	55	80
2	5	310	100	40
3	0	20	200	35
4	0	5	100	465

Fuente: presentación propia de los autores.

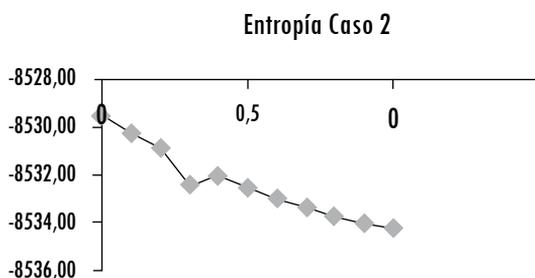
Y se logra la estructura de la solución de entropía analizada en la figura 2.

Es decir, la distribución más probable es aquella que asume los costos máximos y la menos probable se obtiene con la información de la matriz de costos mínimos. Este caso es interesante, pues muestra que el buen análisis de los costos generalizados es un factor importante en las predicciones de viajes urbanos.

Es interesante anotar que en ninguno de los dos casos el costo promedio ($\lambda = 0,5$) generó la máxima entropía; es más, al hacer pequeñas variaciones de dicho valor (por ejemplo, para el caso 2: ($\lambda = 0,6$) se obtienen distribuciones más probables). Por otro lado, dada la no convexidad de las figuras 1 y 2, se puede

concluir que entre más fina sea la malla para obtener λ_k , se conseguirán mejores aproximaciones de la máxima entropía.

Figura 2. Máxima entropía para diferentes valores de l y la matriz a priori 2



Fuente: presentación propia de los autores.

5. Conclusiones y trabajos futuros

En este trabajo se propone una formulación extendida del problema clásico de distribución de viajes con el método de máxima entropía, usando costos intervalares. Para lograr dicha formulación se explica brevemente el problema de la máxima entropía en el contexto probabilístico y su formulación como un problema de distribución de viajes. De esta forma, se logra una formulación robusta, debido a que se sensibilizan a priori los parámetros de los costos, lo cual permite generar un conjunto de soluciones intervalares por la formulación como un problema de elecciones discretas análogo a la solución de modelos con variables aleatorias como parámetros de programación estocástica o problemas de teoría de la decisión. Adicional a la formulación se propone una nueva valoración del parámetro de sensibilidad al costo y las medidas de accesibilidad y atraktividad por la solución que encuentra la máxima entropía.

Un resultado importante de este tipo de modelaciones es que no es necesario tener la información de los costos de viaje, sino estimar el mínimo costo y el máximo, generando análisis de sensibilidad a priori en las soluciones (Wang y Xiao, 2010). Como trabajo futuro se recomiendan varias líneas claras de investigación, como extender esta formulación a otro tipo de funciones de pertenencia difusa. Por otro lado, incluir parámetros difusos en la generación y atracción de viajes para generar mayor robustez al modelo. Sería interesante aplicar la modelación con datos reales para comparar con los resultados clásicos, y debido a la complejidad del modelo, usar otras técnicas de solución, como metaheurísticas y otras formulaciones de entropía, como los costos cuadráticos. Adicionalmente, se podría comparar la solución del problema intervalar con la solución de un

problema asumiendo diferentes distribuciones de probabilidad (programación estocástica) para analizar la robustez del modelo y específicamente de las predicciones. Por otro lado, sería interesante explorar la solución continua de λ resolviendo directamente el problema de optimización de la forma:

$$\begin{aligned} & \max_{T_{ij}, \lambda} - \sum_{i=1}^K \sum_{j=1}^L T_{ij} \ln T_{ij} \\ \text{Sujeto a:} \\ & \sum_{j=1}^L T_{ij} = O_i, \quad i = 1, \dots, K \\ & \sum_{i=1}^K T_{ij} = D_j, \quad j = 1, \dots, L \\ & \sum_{i=1}^K \sum_{j=1}^L (\lambda C_{ij}^{\max} + (1-\lambda) C_{ij}^{\min}) * T_{ij} = \lambda C^{\max} + (1-\lambda) C^{\min} \\ & \lambda \in [0, 1] \end{aligned} \tag{11}$$

Como conclusión general, se tiene que dado que el problema de la maximización de entropía es muy usado en la práctica para estimar y realizar predicciones de distribución de viajes, el modelo presentado genera una solución más robusta, incorporando el hecho de que los costos generalizados de viaje pueden variar (en este caso, en un intervalo con la misma probabilidad para un par origen destino), hasta obtener que la formulación propuesta podría mejorar las predicciones que se realizan en el sector de transporte donde generalmente se usan costos promedios ignorando la variabilidad intrínseca de dicha medida o la poca robustez dada por la presencia de datos atípicos o aglomeraciones por sectores.

El hecho de usar la entropía como criterio de selección permite determinar la distribución más probable dada la información conocida o determinística, como los viajes generados y atraídos por zona y bajo el supuesto de que cada costo es no determinista, y se pueden obtener soluciones diferentes al caso promedio. En las simulaciones se muestra que usar costos promedio genera soluciones menos probables que usando costos asimétricos ($\lambda \neq 0,5$), por lo cual es necesario en la práctica valorar y obtener con bastante cuidado dicho valor, haciendo pertinente la mejor recolección de datos que capturen la variabilidad de los costos para efectuar predicciones y análisis de políticas urbanas y de transporte; obteniendo así la utilidad en términos reales que puede derivarse de esta investigación.

References

- CADENAS, J. y VERDEGAY, J. *Modelos de optimización con datos imprecisos*. Murcia, España: Universidad de Murcia, servicio de publicaciones, 1999.
- DELL'ORCO, M. y KIKUCHI, S. An alternative approach for choice models in Transportation: use of possibility theory for comparison of utilities. *Yugoslav Journal of Operations Research*. 2004, vol. 14, núm. 1, pp. 1-17
- FANG, S.; RAJASEKERA, J. y TSAO, H. *Entropy Optimization and Mathematical Programming*. Norwell, MA: Kluwer Academic Publishers, 1997.
- LÓPEZ, H. *Introducción a GAMS y su aplicación en la solución de modelos matemáticos de optimización. Cuadernillo cursillo XXII Coloquio distrital de Matemáticas y estadística*. Bogotá: Universidad Nacional de Colombia, Departamento de Matemáticas, 2006.
- MARTÍNEZ, F. Access, the Economic Link in Transport-Land Use Interaction. *Transportation Research B*. 1995, vol. 29, núm. 6, pp. 457-47.
- MARTÍNEZ, F. y ARAYA, C. Note on Trip Benefits in Spatial Interaction Models. *Journal of Regional Science*. 2000, vol. 40, núm. 4, pp. 789-796.
- MOMENI, H. Y NAZEMI, J. A fuzzy-random Utility Model of Consumer Choice; Uncertainty Effect. *World Applied Sciences Journal*. 2010, vol. 8, núm. 10, pp. 1183-1192.
- ORTÚZAR, J. y WILLUMSEN, L. *Modelos de transporte*. Santander: Publican, Ediciones de la Universidad de Cantabria, 2008.
- ROMERO, J. *Notas sobre Máxima Entropía. Documento de trabajo*. México: UNAM Facultad de Ciencias, 2008.
- ROMIK, D. Stirling's Approximation for $n!$: The Ultimate Short Proof? *The American Mathematical Monthly*. 2000, vol. 107, núm. 6, pp. 556-557.
- SHANNON, C. E. A mathematical theory of communication. *Bell System Technical Journal*. 1948, vol. 27, pp. 379-423 y 623-656.
- WANG, X. y XIAO, M. Approach of group decision making based on normal distribution interval number with incomplete information. *Journal of Control and Decision*. 2010, vol. 10. 1494-1498.
- WILSON, A. G. *Entropy in urban and regional modeling*. London: Pion, 1970.

