

## SPATIAL INTERACTION MODEL FOR TRIP-CHAINING BEHAVIOR BASED ON ENTROPY MAXIMIZING METHOD

Yudai Honma      Osamu Kurita      Azuma Taguchi  
*Waseda University      Keio University      Chuo University*

(Received November 6, 2009; Revised July 6, 2010)

*Abstract* In this paper, we propose a new spatial interaction model for trip-chaining behavior that consists of a sequence of movements. Particularly, including the origin-destination constraints, we generalize the traditional entropy maximizing model to deal with trip-chaining behaviors. Traditional entropy models should be noted in terms of a theoretical derivation of the gravity model and its validity to real data. However, these models only deal with simple movements from origin to destination. On the contrary, people frequently visit several destinations in one trip and make a sequence of movements. In this regard of view, we extend the traditional entropy model, and propose a general framework for deriving trip-chain distributions incorporating a sequence of movements. This model enables us to estimate the trip-chain distribution to maximize the entropy under several constraints. Finally, we apply the model to a person trip survey in the Tokyo metropolitan area to examine the validity of the model.

**Keywords:** Transportation, spatial interaction model, entropy maximizing method, trip-chain, multiple destinations, Tokyo metropolitan area

### 1. Introduction

In this paper, we propose a new spatial interaction model for trip-chaining behavior that consists of a sequence of movements. In particular, we include the origin-destination constraints and generalize the traditional entropy maximizing model to deal with trip-chaining behaviors. Our main objective is to establish a methodology to estimate the most probable distribution of trip-chains under several constraints.

In previous researches related to spatial interaction, various mathematical models have been developed to reproduce and predict human, commodity, and information flows which are observed in cities. Especially, the entropy maximizing model should be noted in terms of a theoretical derivation of the gravity model and its validity to real data. Entropy models have been proposed by several authors, and all models follow the principle of maximum entropy in physics and mathematics. Although a large number of spatial interaction models have been proposed today, it is not an exaggeration to say that many are triggered by the idea of entropy model.

Traditional spatial interaction models have been developed to estimate a simple movement from origin to destination. However, people frequently visit several destinations in a travel and make a sequence of movements. Behaviors such as “comparing several boutiques to buy clothes” and “visiting multiple sightseeing areas in a journey” can be mentioned as examples. In the field of traffic engineering, each movement from a place to another place is regarded as a “*trip*”, and these sequences of movements are classified as a “*trip-chain*”, One of the features of a trip-chain is that each trip that makes up a behavior would have reciprocal relation.

Trip-chaining behaviors have been discussed in a number of research papers. One of the common methods to analyze them is the Markovian approach. In particular, the series of research works by Sasaki [20–22] is highly important, because the techniques to apply the Markov model to a trip-chain are summarized in detail. Additionally, a variety of other models based on the Markovian approach have also been proposed [5–8, 11, 12]. Meanwhile, various approaches besides the Markovian approach have also been developed. The discrete choice model is another typical method to analyze trip-chaining behavior. These researches basically focus on the formulation of the utility and alternatives of trip-chaining behavior [1–4, 10, 14–16, 23, 25].

However, analyzing trip-chaining behaviors by these two methods have some demerits unfortunately. In particular, there are few foundations that trip-chain behaviors have Markov property. It is regarded as a weak points to apply Markov model to trip-chaining behaviors. There are also some demerits to use discrete choice model. It is essential to formulate utilities of each trip-chain, but trip-chaining behaviors have too much flexibility and it is tough task to identify their utilities. That is to say, both persuasive and coherent methods for trip-chaining behaviors are essential.

Above all, please remind that trip-chaining behavior is a kind of “spatial interaction”, and entropy model can be regarded as a root of spatial interaction models. Therefore entropy maximizing method for trip-chaining behaviors should be also proposed. Nevertheless, only a few researchers approached the trip-chaining behaviors using the entropy maximizing method [13, 18, 26]. Unfortunately, these researches have some restrictions in their formulations, and are difficult to interpret as generalizations of the traditional entropy model. For example, both Tomlinson’s and Mazurkiewicz’s approaches are limited by probable trip-chaining behaviors. In Roy *et al.* by contrast, a variety of behaviors are considered, but there is no constraint for destination (i.e., the number of visitings).

In this regard, we extend the traditional entropy model, and propose a general framework for deriving trip-chain distributions incorporating a sequence of movements. In particular, it is noteworthy that our model incorporates the constraint of the number of visitors in each destination. Since this constraint considers not only the movements from origin zones but also the movements from other destination zones, the complexity of the model is directly dependent on their formulation. We believe that our model can solve this problem, and extend the traditional model simply by intelligible formulation.

This paper is organized as follows. In Section 2, we will discuss the traditional entropy maximizing model. In accordance with Wilson, Sasaki, and Weibull, we summarize the formulation of the doubly-constrained entropy model and the estimation procedure. Then, we discuss the basic idea of our study (Section 3). Our main concerns are what to estimate and how to extend the traditional model. We take a simple case where only 2 zones exist, and describe the concept of the study. In Section 4, an entropy model for trip-chaining behaviors will be presented. This model enables us to estimate the trip-chain distribution to maximize the entropy under both origin-destination constraints and total-transportation-cost constraint. In addition, the procedure to estimate model parameters is discussed. In Section 5, our model is applied to a person trip survey in the Tokyo metropolitan area and the validity of the model is examined.

## 2. Traditional Model

Before representing the main context of the paper, we summarize a traditional entropy maximizing model. This will be helpful for a better understanding of the concept of the

study.

### 2.1. Formulation

The entropy maximizing model (especially, Wilson's entropy model) is often recognized as the root of "the families of spatial interaction models". As is well known, spatial interaction models aim to estimate the flows between origin zones and destination zones. In particular, the entropy model tries to solve this problem based on the law of increase of entropy. The formulation of that entropy model is as follows.

Let us assume origin zones  $i$  ( $i = 1, 2, \dots, I$ ) and destination zones  $j$  ( $j = 1, 2, \dots, J$ ) and define  $T_{ij}$  as the number of trips from  $i$  to  $j$ .  $T_{ij}$  satisfies the following constraint equations:

$$O_i = \sum_{j=1}^J T_{ij} \quad (i = 1, 2, \dots, I), \quad (2.1)$$

$$D_j = \sum_{i=1}^I T_{ij} \quad (j = 1, 2, \dots, J). \quad (2.2)$$

Here,  $O_i$  and  $D_j$  denote the number of individuals leaving origin  $i$  and visiting destination  $j$  respectively. The total number of trips is expressed as  $T$ :

$$T = \sum_{i=1}^I \sum_{j=1}^J T_{ij} \left( = \sum_{i=1}^I O_i = \sum_{j=1}^J D_j \right). \quad (2.3)$$

Furthermore, the following total-transport-cost constraint is imposed:

$$C = \sum_{i=1}^I \sum_{j=1}^J T_{ij} c_{ij}, \quad (2.4)$$

where  $c_{ij}$  is the travel cost from origin  $i$  to destination  $j$  and is predetermined.

The objective of the entropy model is to estimate the most probable trip distribution  $\{T_{ij}\}$  under the assumption that  $O_i$ ,  $D_j$ ,  $C$  and  $c_{ij}$  are already known (i.e., constraints (2.1), (2.2) and (2.4) are satisfied). For this purpose, let us introduce the following function:

$$w(\{T_{ij}\}) = \frac{T!}{\prod_{i=1}^I \prod_{j=1}^J T_{ij}!} \prod_{i=1}^I \prod_{j=1}^J (p_{ij})^{T_{ij}}, \quad (2.5)$$

where  $p_{ij}$  is the prior probability of a trip from  $i$  to  $j$  and is predetermined by analyst. (2.5) follows the idea of multinomial distribution. Prior probability  $p_{ij}$  expresses a priority of one trip from  $i$  to  $j$ , and hence,  $w(\{T_{ij}\})$  is the probability that trip distribution  $\{T_{ij}\}$  is obtained. In addition,

$$\sum_i \sum_j p_{ij} = 1 \quad (2.6)$$

is satisfied. This expression is proposed by several authors (e.g., Wilson [27, 28], Sasaki [19], Snickars and Weibull [24], etc.). In Wilson's formulation, each  $p_{ij}$  is defined equally likely. Some authors, on the contrary, have supposed that  $p_{ij}$  depends on various factors such as the area of zones.

Now, let us maximize (2.5) subject to (2.1), (2.2), and (2.4). Note that it is more convenient to maximize  $\ln w(\{T_{ij}\})$  rather than  $w(\{T_{ij}\})$  itself (this transition has no effect because a logarithmic function is a monotonically increasing function). By using Stirling's approximation  $N! = N \ln N - N$ , we derive

$$\ln w(\{T_{ij}\}) = \ln T! - \sum_{i=1}^I \sum_{j=1}^J (T_{ij} \ln T_{ij} - T_{ij}) + \sum_{i=1}^I \sum_{j=1}^J T_{ij} \ln p_{ij}. \tag{2.7}$$

Thus, the Lagrangian function for this optimization problem is given by

$$\begin{aligned} \mathfrak{L}(\{T_{ij}\}; \lambda, \mu) = & \ln T! - \sum_{i=1}^I \sum_{j=1}^J (T_{ij} \ln T_{ij} - T_{ij}) + \sum_{i=1}^I \sum_{j=1}^J T_{ij} \ln p_{ij} \\ & + \sum_{i=1}^I \lambda_i \left( O_i - \sum_{j=1}^J T_{ij} \right) \\ & + \sum_{j=1}^J \mu_j \left( D_j - \sum_{i=1}^I T_{ij} \right) \\ & + \gamma \left( C - \sum_{i=1}^I \sum_{j=1}^J T_{ij} c_{ij} \right). \end{aligned} \tag{2.8}$$

The  $T_{ij}$ 's which maximize  $\mathfrak{L}$  and that therefore constitute the most probable distribution of trips, are the solutions of

$$\frac{\partial \mathfrak{L}}{\partial T_{ij}} = -\ln T_{ij} + \ln p_{ij} - \lambda_i - \mu_j - \gamma c_{ij} = 0 \tag{2.9}$$

and constraint equations (2.1), (2.2), and (2.4). From (2.9), we derive

$$T_{ij} = p_{ij} \exp[-\lambda_i - \mu_j - \gamma c_{ij}]. \tag{2.10}$$

Furthermore, substituting (2.10) for origin-destination constraints, we get the following equation:

$$\exp[-\lambda_i] \sum_{j=1}^J p_{ij} \exp[-\mu_j - \gamma c_{ij}] = O_i, \tag{2.11}$$

$$\exp[-\mu_j] \sum_{i=1}^I p_{ij} \exp[-\lambda_i - \gamma c_{ij}] = D_j. \tag{2.12}$$

By assuming

$$A_i = \frac{\exp[-\lambda_i]}{O_i}, \tag{2.13}$$

$$B_j = \frac{\exp[-\mu_j]}{D_j}, \tag{2.14}$$

we obtain the final result as follows:

$$T_{ij} = A_i O_i B_j D_j p_{ij} \exp[-\gamma c_{ij}]. \tag{2.15}$$

where, using Equations (2.11)–(2.14), we get

$$A_i = \left( \sum_{j=1}^J B_j D_j p_{ij} \exp[-\gamma c_{ij}] \right)^{-1}, \tag{2.16}$$

$$B_j = \left( \sum_{i=1}^I A_i O_i p_{ij} \exp[-\gamma c_{ij}] \right)^{-1}. \tag{2.17}$$

This is the formulation of the traditional entropy maximizing model. In this study, we generalize the preceding model (2.15)–(2.17) with respect to trip-chaining behavior.

**2.2. Estimating parameters**

As recognized from (2.15), we have to calculate  $A_i$ ,  $B_j$ , and  $\gamma$  to determine trip distribution  $\{T_{ij}\}$ . For this purpose, we state a procedure to estimate the above  $I + J + 1$  parameters under constraints (2.1), (2.2), and (2.4). There are several methods of calibration, and the method we use in the paper is as follows.

First, substituting (2.15) into (2.4), we have

$$f(\gamma) = \sum_{i=1}^I \sum_{j=1}^J A_i O_i B_j D_j p_{ij} \exp[-\gamma c_{ij}] c_{ij} - C = 0. \tag{2.18}$$

Then, using preceding equations (2.16), (2.17), and (2.18), the following iterative algorithm is calculated:

- i) As starting values, we set  $\gamma = \gamma^0$ ,  $B_j = B_j^0$  ( $j = 1, 2, \dots, J$ ), and  $\xi = 0$ .
- ii) Calculate
  - $A_i^{\xi+1} = \left( \sum_{j=1}^J B_j^\xi D_j p_{ij} \exp[-\gamma c_{ij}] \right)^{-1}$  ( $i = 1, 2, \dots, I$ )  
(from (2.16)).
  - Then, calculate
  - $B_j^{\xi+1} = \left( \sum_{i=1}^I A_i^{\xi+1} O_i p_{ij} \exp[-\gamma c_{ij}] \right)^{-1}$  ( $j = 1, 2, \dots, J$ )  
(from (2.17)).
- iii) If  $A_i^{\xi+1} \doteq A_i^\xi$  ( $i = 1, 2, \dots, I$ ) and  $B_j^{\xi+1} \doteq B_j^\xi$  ( $j = 1, 2, \dots, J$ ) are satisfied, then go to iv). If not, set  $\xi = \xi + 1$  and go back to ii).
- iv) Set  $x^0 = \gamma^\xi$  and calculate

$$x^{\kappa+1} = x^\kappa - f(x^\kappa) / f'(x^\kappa)$$

iteratively until  $|x^{\kappa'+1} - x^{\kappa'}| < \varepsilon$  is satisfied, where  $\varepsilon$  is a small positive number. Then, set  $\gamma^{\xi+1} = x^{\kappa'+1}$ .

- v) Set  $\xi = \xi + 1$  and go to ii).

Here, first derivative  $f'(x)$  in iv) is

$$f'(x) = - \sum_{i=1}^I \sum_{j=1}^J A_i O_i B_j D_j p_{ij} \exp[-x c_{ij}] c_{ij}^2. \tag{2.19}$$

### 3. Basic Idea of the Study

In Section 2, we discussed the formulation of the traditional entropy model that is focused on simple flows from one zone to another. The main purpose of our study is to extend the traditional entropy model in terms of trip-chaining behavior. In this section, we describe the basic idea of trip-chaining behavior and prepare for an extended model formulated at length in the next section.

#### 3.1. Definition of a trip-chain

Generally, spatial interaction models estimate movements between two zones. This property is not that serious an issue if we want to estimate simple movements such as commuting behaviors. However, is it enough for activities we experience in daily life?

Let us imagine shopping behavior as an example. When we buy something (especially when it is not groceries), we often compare goods in several different stores. In other words, we visit multiple stores successively as one activity. It means that we often carry out a sequence of movements. This is the main idea of trip-chaining behavior.

Keeping the preceding idea in mind, we first define trip-chain. In this study, we define a *trip-chain* as a sequence of movements which

- (i) starts from an origin zone (indexed by  $i$ ),
- (ii) visits several destination zones (indexed by  $j$ ) successively, and
- (iii) goes back to the same origin zone.

In the previous section, we indexed a trip from origin zone  $i$  to destination zone  $j$  as  $ij$ , and define  $T_{ij}$  as the number of people who make trip  $ij$ . Similarly, we suppose that  $i$  and  $j$  index origin zone and destination zone respectively, and define that a trip-chain  $i\mathbf{j}$  is a series of  $\Lambda + 1$  trips:

$$\text{origin zone } i \rightarrow \text{destination zone } j_1 \rightarrow \dots \rightarrow \text{destination zone } j_\Lambda \rightarrow \text{origin zone } i,$$

where  $\mathbf{j} = [j_1, j_2, \dots, j_\Lambda]$  is a  $\Lambda$ -dimensional vector. In addition, we assume that  $T_{i\mathbf{j}}$  is the number of people who make a trip-chain  $i\mathbf{j}$ . Hence,  $\mathbf{j}$  is interpreted as the “visiting path”, and  $\Lambda$  is different from each trip-chain. Some examples are shown in Figure 1.

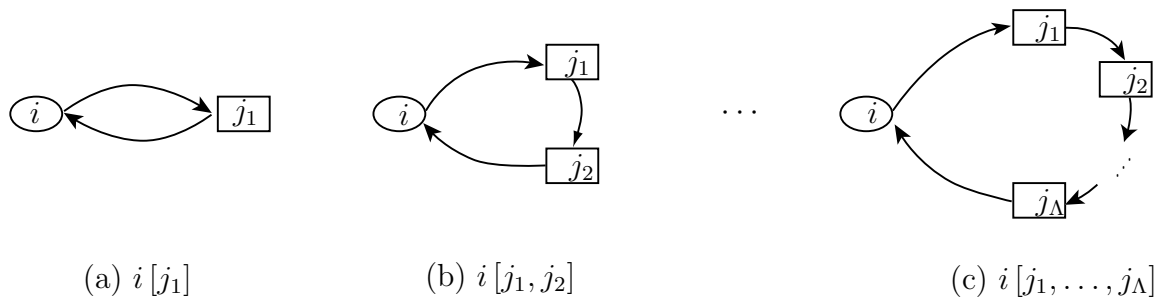


Fig. 1: Examples of trip-chains

#### 3.2. Overview of modeling approach

Suppose that there are two origin zones  $\{i = 1, 2\}$  and two destination zones  $\{j = 1, 2\}$ . As stated in the previous section, spatial interaction models generally estimate trips between origin zones and destination zones ( $2 \times 2 = 4$  cases in this example). As direct generalization, in this study, we estimate  $T_{ij}$  for **all of trip-chaining behavior possible**.

If we allow visit to two destination zones at the most in a trip-chain (i.e.,  $\Lambda \leq 2$ ), then the possible trip-chains in this example are the 12 cases indicated in Figure 2. That is to say,

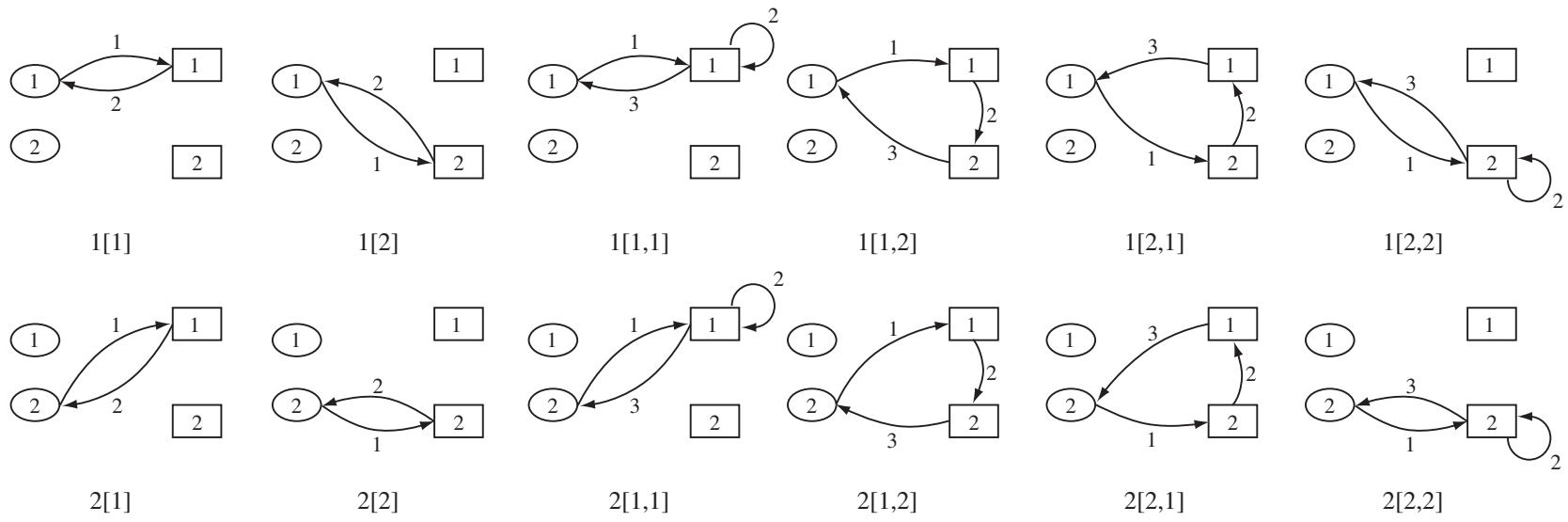


Fig. 2: Possible trip-chain paths ( $I = 2, J = 2, L = 2$ )

- (i) trip-chains that visit one destination zone ( $\Lambda = 1$ )  
origin zones: 2 cases  $\times$  first destination zones: 2 cases = 4 cases,
- (ii) trip-chains that visit two destination zones ( $\Lambda = 2$ )  
origin zones: 2 cases  $\times$  first destination zones: 2 cases  
 $\times$  second destination zones: 2 cases = 8 cases.

In this paper, we regard the 12 trip-chains in Figure 2 as different, and propose an entropy model for trip-chaining behavior. To formulate the most generalized cases, in our study, we incorporate trip-chains that visit the same destination zone successively (e.g.,  $\mathbf{j} = [1, 1]$ ). This assumption seems to be strange, but situations that confirm multiple destinations in the same zone are frequently observed. On the other hand, there are also situations that do not consider some trip-chain paths. We can apply this model to such situations without loss of generality (the method is discussed in detail in Section 4.3).

### 3.3. Extending the model

We now move on to the basic idea of our entropy model for trip-chains. To explain how we can extend the model, we would like to consider a simple example. In particular, we discuss how we can extend the entropy function and constraints for the trip-chains described in Figure 2.

First, we explain how we can generalize entropy function (2.5) to incorporate trip-chaining behavior. Then, we consider the structure of the function. In (2.5), two  $\prod_i \prod_j$  are used and this implies that **all trip paths in the model are multiplied**. In a similar manner, the entropy for trip-chain paths in Figure 2 should be defined as follows:

$$w(\{T_{ij}\}) = \frac{T!}{\underbrace{T_{1[1]}! \cdots T_{1[2,2]}! T_{2[1]}! \cdots T_{2[2,2]}!}_{12 \text{ factors}}} \cdot \underbrace{p_{1[1]}^{T_{1[1]}} \cdots p_{1[2,2]}^{T_{1[2,2]}} p_{2[1]}^{T_{2[1]}} \cdots p_{2[2,2]}^{T_{2[2,2]}}}_{12 \text{ factors}}. \tag{3.1}$$

That is,  $\prod_i \prod_j$  is rewritten to include all trip-chain paths considered in the model.

Next, we state the generalization of constraints. In the traditional model, origin-destination and total-transport-cost constraints are assumed. Hence, we also set these three constraints in the proposed model.

We first describe the origin constraint. As in the traditional model, we define origin-constraints as the number of travelers who depart from each origin zone. We thus get  $O_1$ , the constraint of origin zone 1, as follows:

$$O_1 = T_{1[1]} + T_{1[2]} + T_{1[1,1]} + T_{1[1,2]} + T_{1[2,1]} + T_{1[2,2]}. \tag{3.2}$$

Furthermore,  $O_2$  is calculated as

$$O_2 = T_{2[1]} + T_{2[2]} + T_{2[1,1]} + T_{2[1,2]} + T_{2[2,1]} + T_{2[2,2]}. \tag{3.3}$$

The preceding equations indicate that origin constraints are derived by adding all possible patterns of stops.

We now discuss the constraints of destination zones. In the traditional model, destination constraints express the number of individuals who visit the destination zones. Thus, it is natural that our model also constrain the number of individuals who visit the destination zones. To constrain them, we have to consider the number of individuals who choose  $j$



as first destination and second destination, respectively. For example, the constraint for destination zone  $S_1$  is written as

$$D_1 = \underbrace{T_{1[1]} + T_{2[1]} + T_{1[1,1]} + T_{2[1,1]} + T_{1[1,2]} + T_{2[1,2]}}_{\text{visiting as first destination}} + \underbrace{T_{1[1,1]} + T_{2[1,1]} + T_{1[2,1]} + T_{2[2,1]}}_{\text{visiting as second destination}}. \quad (3.4)$$

Similarly, that for destination zone 2 is

$$D_2 = \underbrace{T_{1[2]} + T_{2[2]} + T_{1[2,1]} + T_{2[2,1]} + T_{1[2,2]} + T_{2[2,2]}}_{\text{visiting as first destination}} + \underbrace{T_{1[1,2]} + T_{2[1,2]} + T_{1[2,2]} + T_{2[2,2]}}_{\text{visiting as second destination}}. \quad (3.5)$$

Finally, we define the total-transport-cost constraint. This is easily described as

$$C = T_{1[1]} \times c_{1[1]} + \dots + T_{1[2,2]} \times c_{1[2,2]} + T_{2[1]} \times c_{2[1]} + \dots + T_{2[2,2]} \times c_{2[2,2]}, \quad (3.6)$$

where  $c_{ij}$  is the transport cost of trip-chain  $ij$  and is predetermined.

Our objective is to maximize entropy (3.1) subject to five constraint equations (3.2)–(3.6), and estimate the number of individuals for each trip-chain in Figure 2. The method for estimation is discussed in detail in the next section.

#### 4. Entropy Model for Trip-Chaining Behavior

In Section 3, we discussed the basic idea of trip-chains and the proposed model. In this section, we state the precise formulation of the entropy model for trip-chaining behavior.

##### 4.1. Formulation

Let  $i$  and  $j$  index origin zones and destination zones respectively ( $i = \{1, 2, \dots, I\}$ ,  $j = \{1, 2, \dots, J\}$ ). In addition, a trip-chain  $ij$  is a series of  $\Lambda + 1$  trips and is defined as

origin zone  $i \rightarrow$  destination zone  $j_1 \rightarrow \dots \rightarrow$  destination zone  $j_\Lambda \rightarrow$  origin zone  $i$ ,

where  $\mathbf{j} = [j_1, j_2, \dots, j_\Lambda]$  is a  $\Lambda$ -dimensional vector. Furthermore, let  $T_{ij}$  be the number of individuals who make trip-chain  $ij$ . For simplicity, we assume that the number of destinations in one trip-chain is less than  $L$  ( $1 \leq \Lambda \leq L$ ).

Our main objective is to estimate  $T_{ij}$  for all trip-chain paths based on the entropy maximizing method. For this purpose, let us suppose the following origin-destination constraints of  $T_{ij}$ :

$$O_i = \sum_{\mathbf{j} \in \Phi} T_{ij} \quad (i = 1, 2, \dots, I), \quad (4.1)$$

$$D_j = \sum_{l=1}^L \sum_{i=1}^I \sum_{\{\mathbf{j} \in \Phi | j_l = j\}} T_{ij} \quad (j = 1, 2, \dots, J), \quad (4.2)$$

where

$$\Phi \stackrel{\text{def.}}{=} [\text{individual's choice set of } \mathbf{j}]. \quad (4.3)$$

For example, in the case of Figure 2,  $\Phi$  is

$$\Phi = \{ [1], [2], [1, 1], [1, 2], [2, 1], [2, 2] \}. \tag{4.4}$$

Furthermore, same as the traditional entropy model, we express the total number of trip-chains as  $T$ :

$$T = \sum_{i=1}^I \sum_{\{j \in \Phi\}} T_{ij} \left( = \sum_{i=1}^I O_i \leq \sum_{j=1}^J D_j \right). \tag{4.5}$$

$\sum_{i=1}^I \sum_{j \in \Phi}$  includes all origin zones and the visiting paths, and thus considers all trip-chain paths in the model. In this model, the summation of  $D_j$  exceeds  $T$ , because individuals have several destinations in one trip-chain path.

Let us explain origin-destination constraints (4.1) and (4.2). Firstly, origin constraint (4.1) is easily obtained by adding all possible visiting paths, and  $\sum_{j \in \Phi}$  in (4.1) expresses this calculation. On the other hand, to derive the destination constraint (4.2), a much more complicated calculation is needed. As stated in Section 3.3, to obtain the total number of individuals who visit destination zone  $j$ , we have to calculate

- “the number of individuals who visit  $j$  as first destination”
- + “the number of individuals who visit  $j$  as second destination”
- +  $\dots$  + “the number of individuals who visit  $j$  as  $L$ th destination”.

For individuals who visit zone  $j$  as  $l$ th destination,  $j_l = j$  is satisfied. This means that

$$\text{“the number of individuals who visit } j \text{ as } l \text{th destination”} = \sum_{i=1}^I \sum_{\{j \in \Phi | j_l = j\}} T_{ij} \tag{4.6}$$

is satisfied. Consequently, by adding (4.6) from  $l = 1$  to  $l = L$ , we obtain the total number of individuals who visit  $j$ , which is the destination constraint of zone  $j$  as (4.2).

Besides the origin-destination constraints, we also assume that  $T_{ij}$  satisfies the total-transport-cost constraint:

$$C = \sum_{i=1}^I \sum_{j \in \Phi} T_{ij} c_{ij}, \tag{4.7}$$

where  $c_{ij}$  is the travel cost of trip-chain  $ij$  per individual and is predetermined.

Let us now estimate the most probable distribution of trip-chains  $\{T_{ij}\}$  under the assumption that  $O_i$ ,  $D_j$ ,  $C$  and  $c_{ij}$  are already known (i.e., constraints (4.1), (4.2) and (4.7) are satisfied). For this purpose, we first derive probability  $w(\{T_{ij}\})$  to find the distribution of trip-chains  $\{T_{ij}\}$ . Now, we assume that  $p_{ij}$  is the prior probability of trip-chain  $ij$ , where

$$\sum_{i=1}^I \sum_{j \in \Phi} p_{ij} = 1, \tag{4.8}$$

and  $p_{ij}$  is predetermined by analyst. Furthermore, as discussed in Section 3.3, probability  $w(\{T_{ij}\})$  is defined by rewriting  $\prod_{i=1}^I \prod_{j=1}^J$  in traditional entropy (2.5), as multiplying all trip-chain paths, that is  $\prod_{i=1}^I \prod_{j \in \Phi}$ . Therefore, we can now derive the probability  $w(T_{ij})$  that obtains a distribution of trip-chains  $\{T_{ij}\}$  as follows:

$$w(\{T_{ij}\}) = \frac{T!}{\prod_{i=1}^I \prod_{j \in \Phi} T_{ij}!} \prod_{i=1}^I \prod_{j \in \Phi} (p_{ij})^{T_{ij}}. \tag{4.9}$$

As in the traditional model, let us maximize probability (4.9) subject to constraints (4.1), (4.2), and (4.7). By using Stirling’s approximation, we derive

$$\begin{aligned} \ln w(\{T_{ij}\}) &= \ln T! - \sum_{i=1}^I \sum_{j \in \Phi} T_{ij}! + \sum_{i=1}^I \sum_{j \in \Phi} (T_{ij} \ln p_{ij}) \\ &\doteq \ln T! - \sum_{i=1}^I \sum_{j \in \Phi} (T_{ij} \ln T_{ij} - T_{ij}) + \sum_{i=1}^I \sum_{j \in \Phi} (T_{ij} \ln p_{ij}). \end{aligned} \tag{4.10}$$

Hence, Lagrangian  $\mathfrak{L}$  is given by

$$\begin{aligned} \mathfrak{L}(\{T_{ij}\}; \lambda, \mu, \gamma) &= \ln T! - \sum_{i=1}^I \sum_{j \in \Phi} (T_{ij} \ln T_{ij} - T_{ij}) + \sum_{i=1}^I \sum_{j \in \Phi} (T_{ij} \ln p_{ij}) \\ &\quad + \sum_{i=1}^I \lambda_i \left( O_i - \sum_{j \in \Phi} T_{ij} \right) \\ &\quad + \sum_{j=1}^J \mu_j \left( D_j - \sum_{l=1}^L \sum_{i=1}^I \sum_{\{j \in \Phi | j_l=j\}} T_{ij} \right) \\ &\quad + \gamma \left( C - \sum_{i=1}^I \sum_{j \in \Phi} T_{ij} c_{ij} \right). \end{aligned} \tag{4.11}$$

The  $T_{ij}$ ’s which maximize  $\mathfrak{L}$ , and that therefore constitute the most probable distribution of trip-chains, are the solutions of

$$\frac{\partial \mathfrak{L}}{\partial T_{ij}} = -\ln T_{ij} + \ln p_{ij} - \lambda_i - \mu_{j_1} - \dots - \mu_{j_\Lambda} - \gamma c_{ij} = 0 \tag{4.12}$$

and constraint equations (4.1), (4.2), and (4.7). This yields the following:

$$T_{ij} = p_{ij} \exp[-\lambda_i - \mu_{j_1} - \dots - \mu_{j_\Lambda} - \gamma c_{ij}]. \tag{4.13}$$

Substituting in equations (4.1) and (4.2) to obtain  $\lambda_i$ , we derive

$$O_i = \exp[-\lambda_i] \sum_{j \in \Phi} p_{ij} \exp[-\mu_{j_1} - \dots - \mu_{j_\Lambda} - \gamma c_{ij}]. \tag{4.14}$$

Doing the same for  $\mu_j$ , we get

$$D_j = \exp[-\mu_j] \sum_{l=1}^L \sum_{i=1}^I \sum_{\{j \in \Phi | j_l=j\}} p_{ij} \exp \left[ -\lambda_i - \mu_{j_1} - \cdots - \underbrace{\mu_{j_{l-1}} - \mu_{j_{l+1}}}_{\mu_{j_l} \text{ does not exist}} - \cdots - \mu_{j_\lambda} - \gamma c_{ij} \right]. \tag{4.15}$$

To obtain the final result in a more familiar form, we denote

$$A_i = \frac{\exp[-\lambda_i]}{O_i}, \tag{4.16}$$

$$B_j = \frac{\exp[-\mu_j]}{D_j}. \tag{4.17}$$

We then obtain

$$T_{ij} = A_i O_i \left( \prod_{l=1}^{\Lambda} B_{j_l} D_{j_l} \right) p_{ij} \exp[-\gamma c_{ij}]. \tag{4.18}$$

Furthermore, substituting (4.18) in (4.1) and (4.2), we derive the following:

$$A_i = \left\{ \sum_{j \in \Phi} \left( \prod_{l=1}^{\Lambda} B_{j_l} D_{j_l} \right) p_{ij} \exp[-\gamma c_{ij}] \right\}^{-1}, \tag{4.19}$$

$$B_k = \left\{ \sum_{l=1}^L \sum_{i=1}^I \sum_{\{j \in \Phi | j_l=j\}} A_i O_i \left( \prod_{\substack{l^*=1 \\ l^* \neq l}}^{\Lambda} B_{j_{l^*}} D_{j_{l^*}} \right) p_{ij} \exp[-\gamma c_{ij}] \right\}^{-1}. \tag{4.20}$$

This is an entropy model for trip-chaining behavior.

#### 4.2. Estimation procedure

Next, we state a procedure to estimate  $T_{ij}$  under constraints (4.1), (4.2), and (4.7). That is to say, our objective is to estimate the spatial interaction, which considers the trip-chain, under the situation that the number of trip from origin zones  $O_i$ , the number of stops in destination zones  $D_j$ , and total-transport-cost in the region is known.

First, substituting (4.18) in (4.7), we have

$$f(\gamma) = \sum_{i=1}^I \sum_{j \in \Phi} A_i O_i \left( \prod_{l=1}^{\Lambda} B_{j_l} D_{j_l} \right) p_{ij} \exp[-\gamma c_{ij}] c_{ij} - C = 0. \tag{4.21}$$

Using the preceding equations (4.19), (4.20), and (4.21), we now make a calibration method that estimates  $I + J + 1$  parameters, namely  $A_i$ ,  $B_j$ , and  $\gamma$ . The algorithm we propose in the paper is as follows:

- i) As starting values, we set  $\gamma = \gamma^0$ ,  $B_j = B_j^0$  ( $j = 1, 2, \dots, J$ ), and  $\xi = 0$ .
- ii) Calculate

$$A_i^{\xi+1} = \left[ \sum_{j \in \Phi} \left\{ \left( \prod_{l=1}^{\Lambda} B_{j_l}^{\xi} D_{j_l} \right) p_{ij} \exp[-\gamma c_{ij}] \right\} \right]^{-1} \quad (i = 1, 2, \dots, I)$$

(from (4.19)).

Then, calculate

$$B_j^{\xi+1} = \left\{ \sum_{i=1}^I \sum_{\{j \in \Phi | j_i=j\}} A_i^{\xi+1} O_i \left( \prod_{\substack{l^*=1 \\ l^* \neq l}}^{\Lambda} B_{j_{l^*}}^{\xi} D_{j_{l^*}} \right) p_{ij} \exp [-\gamma c_{ij}] \right\}^{-1} \quad (j = 1, 2, \dots, J)$$

(from (4.20)).

iii) If  $A_i^{\xi+1} \doteq A_i^{\xi}$  ( $i = 1, 2, \dots, I$ ) and  $B_k^{\xi+1} \doteq B_j^{\xi}$  ( $j = 1, 2, \dots, J$ ) are satisfied, then go to iv).  
If not, set  $\xi = \xi + 1$  and go back to ii).

iv) Set  $x^0 = \gamma^{\kappa}$  and calculate

$$x^{\kappa+1} = x^{\kappa} - f(x^{\kappa}) / f'(x^{\kappa})$$

iteratively until  $|x^{\kappa'+1} - x^{\kappa'}| < \varepsilon$  is satisfied, where  $\varepsilon$  is a small positive number. Then, set  $\gamma^{\xi+1} = x^{\kappa'+1}$ .

v) Set  $\xi = \xi + 1$  and go to ii).

Here, first derivative  $f'(x)$  in iv) is

$$f'(x) = - \sum_{i=1}^I \sum_{j \in \Phi} A_i O_i \left( \prod_{l=1}^{\Lambda} B_{j_l} D_{j_l} \right) p_{ij} \exp [-x c_{ij}] c_{ij}^2. \quad (4.22)$$

### 4.3. Prior probability setting

End of the formulation, we now consider how we can decide the value of prior probability  $p_{ij}$ . In Wilson's formulation [27], it is assumed that each trip is equi-probable. If we follow his idea, prior probability  $p_{ij}$  should be defined as equally as possible. However, we have to be cautious that the number of visiting zones in each trip-chain is different. In other words, simple trip-chains visit only one zone and complicated trip-chains visit multiple zones. Therefore, in our study, we suppose that the prior probability of a trip-chain depends on the number of visitings, and set  $p_{ij}$  as follows:

$$p_{ij} \propto g(\Lambda), \quad (4.23)$$

where  $g(\Lambda)$  is a function which depends on the number of visiting  $\Lambda$ .

As  $p_{ij}$  is directly linked to  $T_{ij}$  by (4.18), we are also able to use prior probability  $p_{ij}$  to apply the model to a more complicated situation. As an example of such a situation, we explain how we can eliminate some trip-chain paths from the model. As confirmed in Figure 2, the formulation stated above incorporates the trip-chains that visit the same destination zone cumulatively. Meanwhile, we should not consider trip-chains that visit a destination zone several times in some situations. Furthermore, it might be appropriate to assume that all people visit destination zones to minimize travel costs (i.e., following the solution of TSP). It is certain that we can incorporate the preceding ideas by setting the prior probability as follows:

$$p_{ij} \stackrel{\text{def.}}{=} 0 \quad (\text{if trip-chain } ij \text{ should not be considered in the model}). \quad (4.24)$$

This parameter setting can be easily implemented because it is calculated only once before iterative estimation.

## 5. Application to Tokyo Metropolitan Area

In this section, we apply the model in Section 4 to the person trip data to examine the validity of the model. We adopt the “JR East Japan Station-consumption-data” as real transportation data in Tokyo, and compare the real flow and estimated flow in detail. Specifically, we pay attention to the features and problems of the model in terms of (i) transportation cost, (ii) flows between zones, and (iii) the number of visiting zones.

### 5.1. Applied data

First, we summarize the JR East Japan Station-consumption-data, which is applied to the model. This is a person trip data originally surveyed by JR East Japan Company. It covers 10013 individuals in the Tokyo metropolitan area, and summarized both movements and consumption activities of each individual from July 5, 2002 to July 8, 2002.

Since, the original purpose of this survey is to analyze the consumption activities of individuals, the preceding data clarifies many items related to their activities. For example, we can identify the (i) date, (ii) stations they use, (iii) their activities, (iv) time of departure and places visited, from the data. Individuals often visit several places in one trip-chain, and (ii)–(iv) are recorded for each visiting. In addition, this data records all movements from departure to return, so each movement is regarded as a trip-chain whose origin is the individual’s houses. In this paper, we used the “movement part of data” as transportation data of the Tokyo metropolitan area (the “consumption activity part of data” is not utilized because it is private information).

### 5.2. Constraint setting

Next, we summarize how we apply the data to the model. To begin with, we explain the assumption of zones. As stated above, since we figure out their movements based on stations, we could regard each station as a different zone. However, there are about 1800 stations in the data, and a large number of visiting paths must be considered if each station is regarded as a zone. In this study therefore, we set 5 km grids as indicated in Figure 3, and regard the data as movements between these grids. As a result, we use 10831 samples, wherein movement is within 60 km from the Imperial Palace and the number of visitings is less than three (i.e.,  $\Lambda \leq 3$ ).

Based on the grids assumed above, we calculated the constraints of the model. We assume that each grid is both the origin zone and the destination zones, and calculate the origin-destination constraints. The distribution of constraints is demonstrated in Figure 4. In addition, the total-transport-cost constraint is calculated from the real data. Here, the transportation cost between two zones is calculated as the Euclidean distance that incorporates a detour of Tokyo Bay, and the transportation cost of a trip-chain is assumed to be the summation of each trip cost.

Finally, we discuss how we set the prior probabilities. As previously mentioned in Section 4.3, we have to determine the prior probabilities before estimation. In this study, to consider the most general case, we assume that the all prior probabilities  $\{p_{ij}\}$  are constant. Of course, we can suppose more sophisticated prior probabilities, especially when we have enough information. However, if there is no prior information about the real flow, then it is difficult to set such prior probabilities. To suppose the worst case, we assume that all trip-chains are equi-probable.

### 5.3. Estimation

Using the model proposed in Section 4, we estimated trip-chain distribution  $\{T_{ij}\}$  under the preceding constraints. We compare the estimated flow with the real flow and evaluate the validity of our model from several points of view.

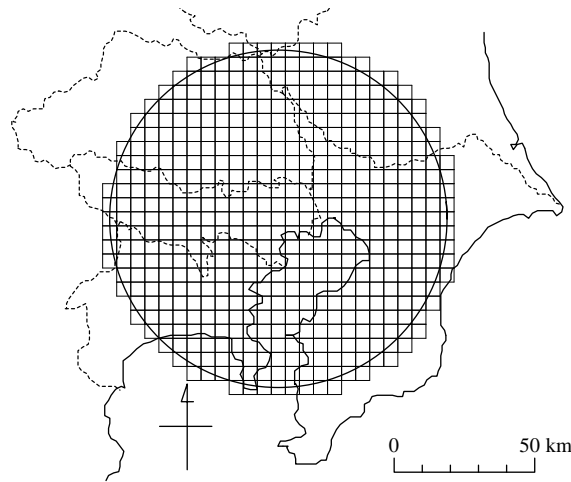


Fig. 3: Applied region and 5 km grids

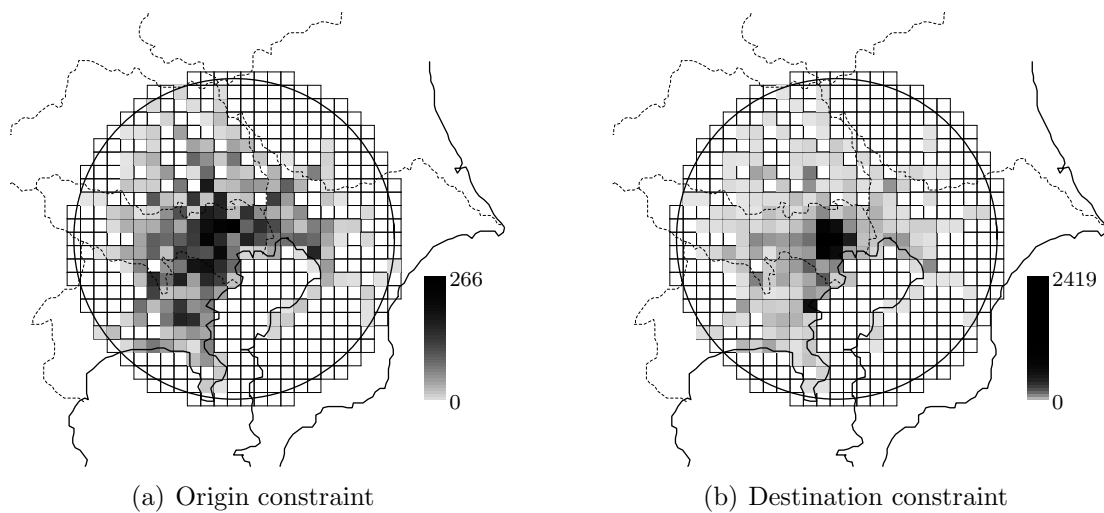


Fig. 4: Distribution of constraints

It is important to adopt evaluation indices appropriately. Here let us remind that the purpose of our model is to estimate the number of individuals for each trip-chain path  $\{T_{ij}\}$ . Thus, it is meaningful to compare the real value of  $T_{ij}$  and the model value. However, these values generally take low value because there are so many candidate trip-chain paths in regions. Hence, it is also favorable to compare the aggregated value of  $T_{ij}$ . In particular, trip-chaining behavior can be categorized such as (i) the length of travel cost, (ii) each movements which makes trip-chain, or (iii) the number of visits. We select the evaluation indices in terms of the importance in other analyses, and evaluate not only  $T_{ij}$  itself but also the aggregated value of  $T_{ij}$ .

### 5.3.1. Distribution of transportation cost

First, we compare the transportation cost per trip-chain  $c_{ij}$ . In particular, we calculated the distribution of transportation cost, which is frequently used to analyze the flows in regions [9]. Figure 5 illustrates the aggregation of the number of individuals for every 10 km trip-chain about both the real flows and the estimated flows. The horizontal axis indicates the transportation cost per trip-chain, and the vertical axis is the number of

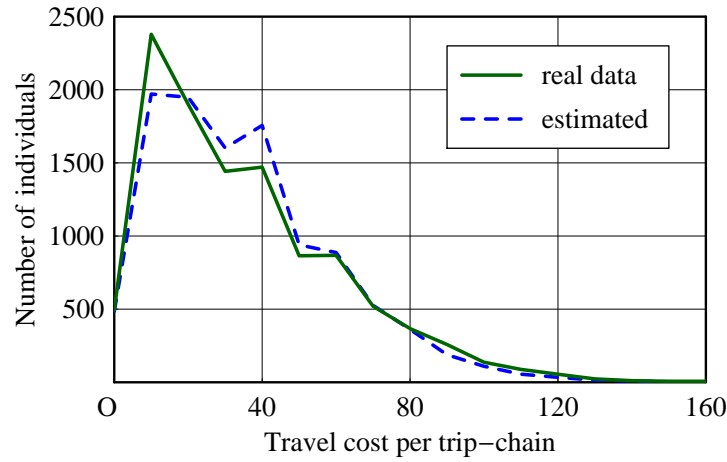


Fig. 5: Distribution of travel cost

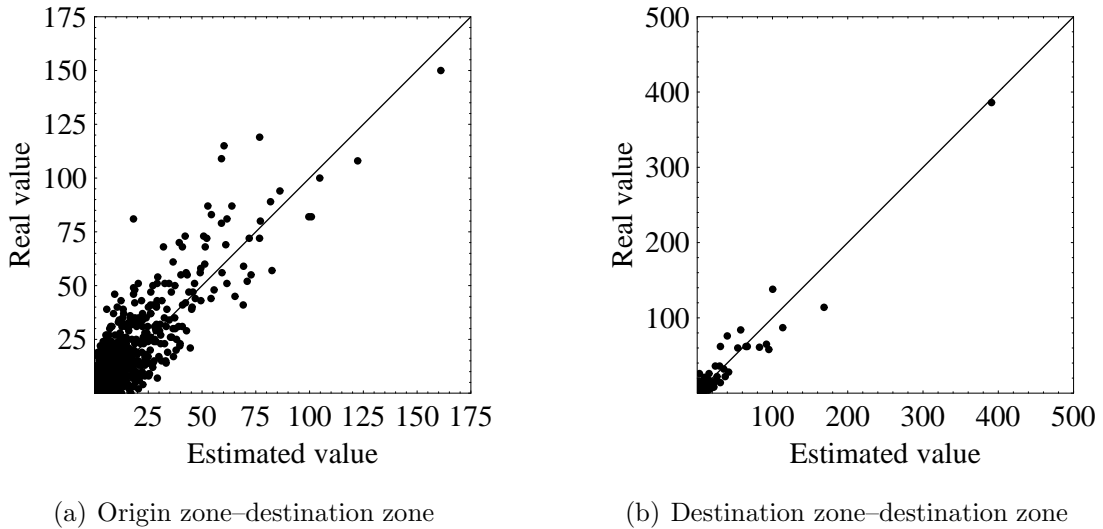


Fig. 6: Comparison of flows between zones

individuals. Though the distribution when travel cost is less than 40 km is slightly different, we can confirm that both distributions fit well enough. This feature results from the total-transportation-cost constraint of the model.

**5.3.2. Number of flows between two zones**

Next, to analyze more deeply, we compare the number of flows between two zones; (i) flows between the origin zone and the destination zone and (ii) flows between two destination zones. As analyzed in lots of studies, flows between two zones are quite important because they indicate the strength of relationship between zones. Figure 6 demonstrates these; the horizontal axes are the estimated value and the vertical axes are the real value. From Figure 6, we know that both flows are relatively well reproduced. The traditional models only aimed to reproduce the flow between origin-destination zones. On the other hand, our model incorporates the idea of a trip-chain, and we calculate not only the flow between origin-destination zones but also the flow between two destination zones. Consequently, it is considerable that the flows between two destination zones fit well because this is one of



Table 1: Number of individuals for each visiting

	1 zone	2 zones	3 zones
Real data	7669	2577	585
Estimated data	7883.23	2149.00	799.13

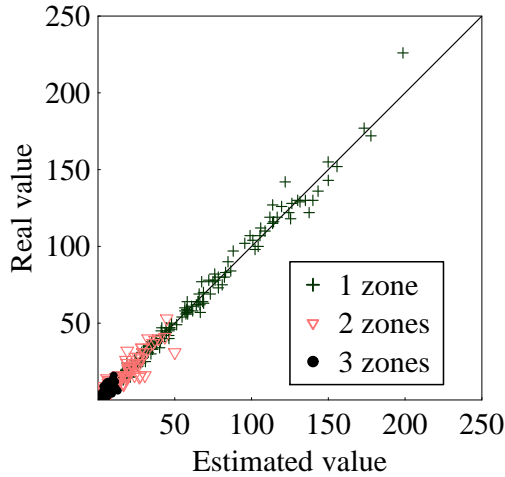


Fig. 7: Comparison of the number of individuals for each origin zone and stop

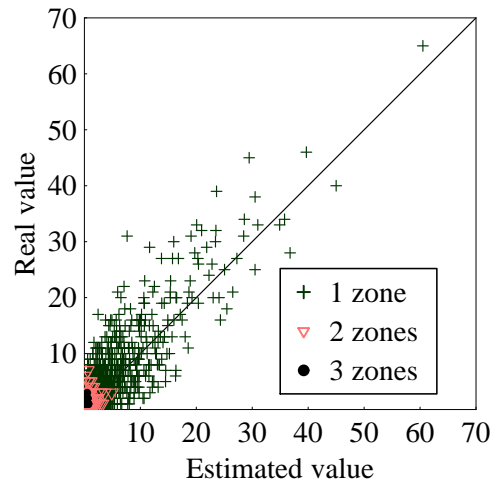


Fig. 8: Comparison of the number of individuals for each trip-chain

the most important points of the study.

**5.3.3. Number of zones visited in each trip-chain**

Now, let us move on to the number of zones visited in each trip-chain. To begin with, we compare the number of visitings for every stops (Table 1). Multiple stops are features of trip-chaining behavior, so it is favorable that the above index is reproduced well. From the table, we confirm that the number of individuals who visit 2 zones is underestimated and that of those who visit 3 zones is overestimated. However, the differences are not that big and the model reproduces relatively well. Generally, the discussion in Section 4.3 is to adjust the fitness of Table 1 (please note that  $p_{ij} \propto f(\Lambda)$  is suggested). Therefore, we consider that it is not so difficult to improve the fitness of Table 1 if necessary.

**5.3.4. Number of individuals for every origin and stop**

To examine the model more precisely, we also calculate the number of individuals for every origin and stop about real data and estimated data respectively (Figure 7). The horizontal axis indicates the estimated value and the vertical axis indicates the real value. Furthermore, the difference in the plot shapes corresponds to the number of stops. Since we derive the model under the origin constraints, it is certain that the number of trips corresponds quite well.

**5.3.5. Number of individuals for each trip-chain**

Finally, we compare the number of trip-chains itself. That is to say, we directly evaluate the distribution of  $\{T_{ij}\}$ , which is estimated in our model. Figure 8 represents the number of individuals for each trip-chain. As in Figure 8, the horizontal axis indicates the estimated

value, the vertical axis indicates the real value, and the plot shapes are the differences in stops. Since the number of individuals is not that high, it is difficult to argue the number of individuals for multi-stops (in fact, real values for many 3 stops are 1 or 2). To compare in terms of the preceding condition, we need more samples of person-trip data. Meanwhile, the number of individuals for a 1-stop trip-chain is generally reproduced well.

### 5.3.6. Adjustment by the prior probabilities

As described from the preceding analysis, the model generally reproduce the real data well. In addition, we can expect to derive a better estimation by adjusting the transportation cost and prior probabilities. For example, differences indicated in Table 1 can be easily adjusted by setting  $p_{ij}$  like (4.23). Furthermore, if  $p_{ij}$  is allowed to depend on travel cost, the fitness of Figure 5 and other results will be much improved. When treating time-series data, it is also possible to estimate the prior probabilities from previous data. We consider that these adjustment enables us to incorporate the characteristics of regions.

Above all, what we should emphasize is that our model keeps the above fitness nevertheless the prior probabilities are set equi-probable as the worst case. Consequently, we conclude that our model has the capability to estimate the real flow with respect to trip-chaining behavior.

## 6. Conclusion

In this paper, we generalized the traditional doubly-constrained entropy model, and proposed a new entropy model for trip-chaining behavior. This model enables us to estimate the most probable trip-chain pattern with respect to origin-destination and total-transportation-cost constraints. Furthermore, we applied the model to a person trip survey in the Tokyo metropolitan area, and examined the validity of the model.

Though we stated the role of prior probability in the paper, how the prior probability affects the trip-chain pattern is not discussed enough. Thus, it should be discussed the theoretical relationship between various types of prior probability and the trip-chain pattern. Further, considering an appropriate setting of the prior probability is an important task that we leave for the future.

## References

- [1] T. Adler and M. Ben-Akiva: A theoretical and empirical model of trip chaining behavior. *Transportation Research*, **13B** (1979), 243–257.
- [2] J. Bowman and M. Ben-Akiva: Activity based disaggregate travel demand model system with daily activity schedules. *Transportation Research*, **35A** (2001), 1–28.
- [3] J. Horowitz: A utility maximizing of the demand for multi-destination non-work travel. *Transportation Research*, **14B** (1980), 369–386.
- [4] J. Horowitz: Modeling traveler responses to alternative gasoline allocation plans. *Transportation Research*, **16A** (1982), 117–133.
- [5] F.E. Horton and P.W. Shuldiner: The analysis of land-use linkages. *Highway Research Record*, **165** (1967), 96–107.
- [6] F.E. Horton and W.E. Wagner: A Markovian analysis of urban travel behavior: Pattern responses by socioeconomic-occupational groups. *Highway Research Record*, **283** (1969), 19–29.
- [7] R. Kitamura and T. Lam: A time dependent Markov renewal model of trip chaining. *Proceedings of the 8th International Symposium on Transport and Traffic Theory*,

- (1983), 376–402.
- [8] K. Kondo: Estimation of person trip patterns and modal split. *Transportation and Traffic Theory*, (1974), 715–742.
- [9] T. Koshizuka: Distance distribution in an arbitrary region. *Journal of the City Planning Institute of Japan*, **37** (2002), 37–42. (in Japanese)
- [10] K. Kuroda, S. Yamashita, and F. Akakura: Developing sightseeing excursion model subject to time constraints and estimating the value of improvements road infrastructure. *Proceedings of Infrastructure Planning*, **16** (1993), 293–298. (in Japanese)
- [11] S.R. Lerman: The use of disaggregate choice models in semi-Markov process models of trip chaining behavior. *Transportation Science*, **13** (1979), 273–291.
- [12] D.F. Marble: A simple Markovian model of trip structures in a metropolitan region. *Regional Science Association Western Section Papers*, (Urban System Committee of Arizona State University, Tempe, 1964), 150–156.
- [13] L. Mazurkiewicz: A statistical model of a multitrip spatial-interaction pattern. *Environment and Planning A*, **17** (1985), 1533–1539.
- [14] S. Mizokami, H. Morisugi, and M. Fujita: Modelling on the attraction of sightseeing area and excursion behaviour. *Journal of the City Planning Institute of Japan*, **27** (1992), 517–522. (in Japanese)
- [15] T. Morikawa, K. Sasaki, and R. Azuma: Modeling sightseeing travel behavior for evaluation of road network improvement in the recreational area. *Infrastructure Planning Review*, **12** (1995), 539–547. (in Japanese)
- [16] H. Morisugi, Y. Hayashiyama, and K. Hirayama: Sightseeing behavior forecast by aggregate nested logit-model. *Proceedings of Infrastructure Planning*, **8** (1986), 353–358. (in Japanese)
- [17] J. R. Roy: *Spatial Interaction Modelling—A Regional Science Context* (Springer, Berlin, 2004).
- [18] J. R. Roy, N.C. Smith, and B. Xu: Simultaneous modelling of multi-purpose/multi-stop activity patterns and quantities consumed. *Geographical Systems*, **3** (2001), 303–324.
- [19] T. Sasaki: Trip no OD bunpu wo motomeru kakuritsuron-teki hoho. *Journal of Japan Society of Traffic Engineers*, **2** (1967), 12–21. (in Japanese)
- [20] T. Sasaki: Person trip pattern no 1 bunrui-ho. *Journal of Japan Society of Traffic Engineers*, **4** (1969), 3–9. (in Japanese)
- [21] T. Sasaki: Estimation of person trip patterns through Markov chains. *Traffic Flow and Transportation* (American Elsevier, New York, 1971), 119–130.
- [22] T. Sasaki and Y. Iida: *Kotsu-Kogagu* (Ohmsha, Tokyo, 1992). (in Japanese)
- [23] K. Sasaki, T. Tamura, Y. Masuya, and K. Saito: Application of neural network model to sightseeing travel behavior. *Journal of the City Planning Institute of Japan*, **31** (1996), 367–372. (in Japanese)
- [24] F. Snickars and J.W. Weibull: A minimum information principle: Theory and practice. *Regional Science and Urban Economics*, **7** (1977), 137–168.
- [25] Y. Sugie, A. Fujiwara, M. Moriyama, M. Okumura, and J. Zhang: Influence of road improvement on touring behavior. *Infrastructure Planning Review*, **16** (1999), 699–705.
- [26] J. Tomlinson, N. Bullock, P. Dickens, P. Steadman, and E. Taylor: A model of students' daily activity patterns. *Environment and Planning*, **5** (1973), 231–266.
- [27] A.G. Wilson: A statistical theory of spatial distribution models. *Transportation Research*, **1** (1967), 253–269.

[28] A.G. Wilson: *Entropy in Urban and Regional Modelling* (Pion, London, 1970).

Yudai Honma  
Waseda Institute for Advanced Study  
Waseda University  
3-4-1 Okubo, Shinjuku  
Tokyo 169-8555, Japan  
E-mail: yudai@aoni.waseda.jp