


Article

Optimization in Item Delivery as Risk Management: Multinomial Case Using the New Method of Statistical Inference for Online Decision

Sapto Wahyu Indratno^{1,2}, Kurnia Novita Sari² and Mokhammad Ridwan Yudhanegara^{3,*} 

- ¹ University Center of Excellence on Artificial Intelligence for Vision, Institut Teknologi Bandung, Natural Language Processing & Big Data Analytics (U-CoE AI-VLB), Bandung 40132, West Java, Indonesia; sapto@math.itb.ac.id
- ² Statistics Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung 40132, West Java, Indonesia; kurnia@math.itb.ac.id
- ³ Mathematics Education Department, Faculty of Teacher Training and Education, Universitas Singaperbangsa Karawang, Karawang 41361, West Java, Indonesia
- * Correspondence: mridwan.yudhanegara@staff.unsika.ac.id; Tel.: +62-817-228-396

Abstract: Online activity increasing spreads with the power of technological development. Many studies reported the impact of online activities on decision making. From the statistical perspective, decision making is related to statistical inference. In this regard, it is interesting to propose a new method of statistical inference for online decisions. This method is built by the logarithm distribution of the likelihood function, which allows us to determine statistics using the normal statistical test approach iteratively. It means that the inference can be made in an online way every time new data arrive. Compared to classical methods (commonly, chi-squared), the advantage of this method is that it allows us to make decisions without storing large data. In particular, the novelty of this research is expressed in the algorithm, theorem, and corollary for the statistical inference procedure. In detail, this paper's simulation discusses online statistical tests for multinomial cases and applies them to transportation data for item delivery, namely traffic density. Changes in traffic density resulted in changes to the strategy of item delivery. The goal is to obtain a minimum delivery time for the risk of losses.

Keywords: statistics test; optimization in risk management; online decision; dynamic network



Citation: Indratno, Sapto Wahyu, Kurnia Novita Sari, and Mokhammad Ridwan Yudhanegara. 2022. Optimization in Item Delivery as Risk Management: Multinomial Case Using the New Method of Statistical Inference for Online Decision. *Risks* 10: 122. <https://doi.org/10.3390/risks10060122>

Academic Editors:
Mogens Steffensen and
Krzysztof Jajuga

Received: 15 April 2022
Accepted: 4 June 2022
Published: 10 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

This paper is a continuation of research from Yudhanegara et al. (2021a) on network clustering using item delivery strategies and predictive distribution, as well as Yudhanegara et al. (2021b) on the importance of network clustering, which affects the optimization of item delivery time, and Yudhanegara et al. (2022), which proposed a new strategy in the item delivery process. The focus of the problem presented in this paper is a unique alternative related to statistical inference that is needed in online decision making.

The results of the statistical inference become the basis for carrying out an item delivery strategy, which includes the network clustering process and time optimization. The method formulated in this study provides an efficient theoretical contribution in dealing with the challenges of extensive data and the dynamics of incoming information.

The knowledge required is the involvement of statistics and graph theory (network) in solving research problems. This research problem is raised from the latest natural phenomenon: the item delivery system, which involves the traffic system, time, and distance traveled. The role of statistics in answering real-time data (online) is packaged by utilizing the Bayesian method to determine predictive distributions and statistical inference in decision making. In addition, the multinomial distribution, which has the parameter Dirichlet distribution, is briefly examined. The distribution is considered the

most representative distribution in the item delivery process in a traffic system with many path choices. As for the construction of the network and its clustering, graph theory can be used.

The subject of this research consists of four parts: network clustering, optimization, predictive distribution, and new statistical inference procedures for online decision making. The ultimate goal of integrating the four subjects is to produce an item delivery strategy that can overcome dynamic network conditions. This is important because traffic density changes over time, especially in urban areas or densely populated areas. Furthermore, optimization in item delivery can be used as risk management for goods delivery companies.

We need a precise and efficient statistical inference method. The contribution of this paper is to provide a new alternative for inferential statistics required to make decisions online. Thus, we can make inferences with every arrival of further information or data. The advantage of the method, i.e., it does not require historical data storage, which requires low memory. We make inferences using the previous results. Unlike the classic method (Oosterhoff and Van Zwet 1972; Owen 2018), we need historical data storage, which requires a large memory.

The type of data distribution discussed in this paper is assumed to have a multinomial distribution. Classically, we use chi-squared for the multinomial goodness-of-fit test (Azen and Walker 2010; Walpole et al. 2012). This paper promotes the new method for the multinomial goodness-of-fit test. The method uses a normal distribution as inferential statistics for online decisions. Thus, the hypothesized assumptions for model fit use the Z statistical test. In the future, there will be many issues regarding big data, so that the role of statistics using classical methods is no longer efficient. The new method of inferential statistics for online decisions in this era of big data gives statistics that are still needed.

The new method aims to help practitioners, researchers, and users solve problems in cases of the multinomial distribution. The characteristics of the multinomial distribution, i.e., each experiment has more than two possible events to occur; each experiment is statistically independent so that the events resulting from one experiment do not affect subsequent trials; and the probability of each event in each experiment does not change. The outline of the multinomial discusses data with the number of frequencies in a data category. These multinomial cases have been widely used by practitioners, researchers, and users covering various disciplines, such as politics (Nownes 1992), social (Brock and Durlauf 2003), engineering (Rith et al. 2019), transportation (Reddy et al. 2015), medicine (Nelissen et al. 2020), and other exact sciences. For this reason, the presence of this new method is expected to add new references.

The simulation of the number of vehicles on each road segment is assumed to have a multinomial distribution. Suppose we represent a road map as a network—traffic conditions with the number of vehicles changing every time cause network conditions to be not static. Hence, we need a distribution to describe the dynamic situation of the network. Thus, the predictive distribution function and the estimation of the required parameters are of great importance. For prediction, the predictive distribution approach is used. Predictive distribution is appropriate for cases with known data distribution (Klugman et al. 2012).

In general, the Section 1 (Introduction) of this paper describes the importance of the methods proposed in this paper. The Section 2 (Literature Review) explains the literature on optimization in risk management. The theoretical study of predictive distribution, prediction accuracy, network clustering, and item delivery strategy are described in the Section 3 (Material and Methods). The new methods, limitations, simulation, and result optimization as risk management are discussed in the Section 4 (Result and Discussion). The conclusions are described in the Section 5 (Conclusions).

2. Literature Review

Mathematical optimization, or optimization, is part of operations research. Optimization is a way to determine suitable solutions by minimizing or maximizing one or more objective functions that satisfy all constraints. Optimization is widely used in industrial engineering. Examples of optimization methods include traveling salesman problems, vehicle routing problems, etc.

Optimization is also part of risk management (Aranburu et al. 2016). Risk management is an iterative process and assists in setting strategy, achieving goals, and making informed decisions (Degtereva et al. 2022). Decision making and optimization under uncertainty constitute a broad and popular area of operations research and management sciences (Krokhmal et al. 2011).

The role of risk management is essential for mathematical optimization under uncertainty. Whenever uncertainty exists, there is a risk. Uncertainty is present when there is a possibility that the outcome of a particular event will deviate from what is expected (Better et al. 2008).

There are not many studies that examine optimization in risk management until now. One of the old studies that examine optimization in risk management is in civil engineering conducted by Cooke and Pinter (1989). The information that can be obtained from this research, i.e., the optimization concept and technique can be combined with probabilistic analysis to structure and solve risk management problems. Other research on optimization in risk management can be seen in the study of Better et al. (2008); Krokhmal et al. (2011); Aranburu et al. (2016); and Lam (2016).

Based on the literature study, optimization in Item Delivery as Risk Management has not been found. For this reason, this study is one part of the contents of this paper. This research resulted in how to make online decisions on item delivery strategies to reduce the risk of loss. These losses can be identified through changes in optimizing the delivery time obtained based on the dynamic network.

3. Material and Methods

3.1. Predictive Distribution

Another term for the predictive distribution is the posterior predictive distribution (Klugman et al. 2012). In general, the predictive distribution is the distribution of the possible unobserved values conditional on the observed value. In this case, we will discuss the predictive distribution of the multinomial case. Of course, it involves the Dirichlet distribution because it is prior for the multinomial. The multinomial distribution is denoted $\text{Mult}(\theta_1, \theta_2, \dots, \theta_m; n)$, with the probability mass function as follows:

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{n!}{\prod_{i=1}^m x_i!} \prod_{i=1}^m \theta_i^{x_i}, \quad (1)$$

where $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$, $\sum_{i=1}^m x_i = n$, $n \in \mathbb{N}$, $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_m]^T$, $\sum_{i=1}^m \theta_i = 1$, $\theta_i > 0$ and $\theta_1, \theta_2, \dots, \theta_m$ is the probability parameter. The expected value of the multinomial distribution, namely:

$$E[\mathbf{X}] = [E[X_1], E[X_2], \dots, E[X_m]]^T = [n\theta_1, n\theta_2, \dots, n\theta_m]^T. \quad (2)$$

Then, the Dirichlet distribution is denoted $\text{Dir}(\alpha_1, \alpha_2, \dots, \alpha_m)$, with the probability density function as follows:

$$f(\boldsymbol{\theta}; \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{i=1}^m \alpha_i)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^m \theta_i^{\alpha_i-1}, \quad (3)$$

where $\alpha = [\alpha_1 \alpha_2 \dots \alpha_m]^T$, $\Gamma(\alpha) = \int_0^\infty \theta^{\alpha-1} e^{-\theta} d\theta$, $m \geq 2$, and $\alpha_1, \alpha_2, \dots, \alpha_m$ is the concentration parameter, $\alpha_i > 0$. The expected value of the Dirichlet distribution, namely:

$$E[\Theta] = [E[\theta_1], E[\theta_2], \dots, E[\theta_m]]^T = \left[\frac{\alpha_1}{\sum_{i=1}^m \alpha_i}, \frac{\alpha_2}{\sum_{i=1}^m \alpha_i}, \dots, \frac{\alpha_m}{\sum_{i=1}^m \alpha_i} \right]^T. \tag{4}$$

Let $\mathbf{X}_t = [X_{1,t} X_{2,t} \dots X_{m,t}]^T$ be a random vector that stating the number of events at t^{th} time, where $X_{i,t}$ is a random variable stating the number of events at category i at t^{th} time. Then, assume \mathbf{X}_t follows a multinomial distribution with the parameter θ . Next, we will find the conditional probability distribution for the \mathbf{x}_{t+k} observation if given $\mathbf{D} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{t+(k-1)})$:

$$p(\mathbf{x}_{t+k} | \mathbf{D}_{t+(k-1)}) = \int p(\mathbf{x}_{t+k} | \theta) f(\theta | \mathbf{D}) d\theta. \tag{5}$$

Thus, we obtain Equation (6) (Yudhanegara et al. 2021a):

$$p(\mathbf{x}_{t+k} | \mathbf{D}_{t+(k-1)}) = \frac{n! \Gamma(\sum_{i=1}^m \alpha'_i)}{\Gamma(n + \sum_{i=1}^m \alpha'_i)} \prod_{i=1}^m \frac{\Gamma(x_i + \alpha'_i)}{x_i! \Gamma(\alpha'_i)}. \tag{6}$$

Equation (6) is a predictive distribution probability function which is a probability mass function of the Dirichlet-multinomial (Avetisyan and Fox 2012). The expected value of Equation (6) can be found by utilizing the multinomial expected value and Theorem 1.

Theorem 1. (Hogg et al. 2005). Let Y and X be random variables, then $E[Y] = E[E[Y|X]]$.

Based on Theorem 1, we can find $E[\mathbf{X}_{t+k}] = E[E[\mathbf{X}_{t+k} | \Theta]]$. In the other, we have $(\mathbf{X}_{t+k} | \Theta) \sim \text{Mult}(\theta_1, \theta_2, \dots, \theta_m, n)$; thus:

$$E[\mathbf{X}_{t+k}] = nE[\Theta] = n \left[\frac{\alpha_1}{\alpha_0}, \frac{\alpha_2}{\alpha_0}, \dots, \frac{\alpha_m}{\alpha_0} \right]^T. \tag{7}$$

In the Dirichlet-multinomial case, we can take the value $\alpha_0 = \sum_{i=1}^m \alpha_i$ dan $\theta_i = \frac{\alpha_i}{\sum_{i=1}^m \alpha_i} = \frac{\alpha_i}{\alpha_0}$ (Johnson et al. 1996), so that the following relationship can be seen:

$$E[X_{i,t+k}] = n \theta_i = n \frac{\alpha_i}{\alpha_0}. \tag{8}$$

Thus, we can find θ_i from the multinomial estimated parameter, namely:

$$\hat{\theta}_i = \frac{\sum_{t=1}^l x_{i,t}}{l n}, \quad i = 1, 2, \dots, m, \tag{9}$$

where $\sum_{i=1}^m x_i = n$, $n \in \mathbb{N}$.

3.2. Prediction Accuracy

There are several calculations for prediction errors, one of which is the mean absolute percentage error (MAPE), which is the absolute average percentage error. MAPE is a statistical measure of predictive accuracy in predictive models that provides information on how much prediction error is compared to the observed value based on the sequence (Swamidass 2000). According to Lewis (1982), MAPE scores can be interpreted into four criteria. There are predetermined criteria as described in Table 1.

Table 1. Criteria for the MAPE score.

Percentage	Criteria
MAPE < 10%	Very accurate
10% ≤ MAPE < 20%	Good
20% ≤ MAPE < 50%	Fair
MAPE ≥ 50%	Not accurate

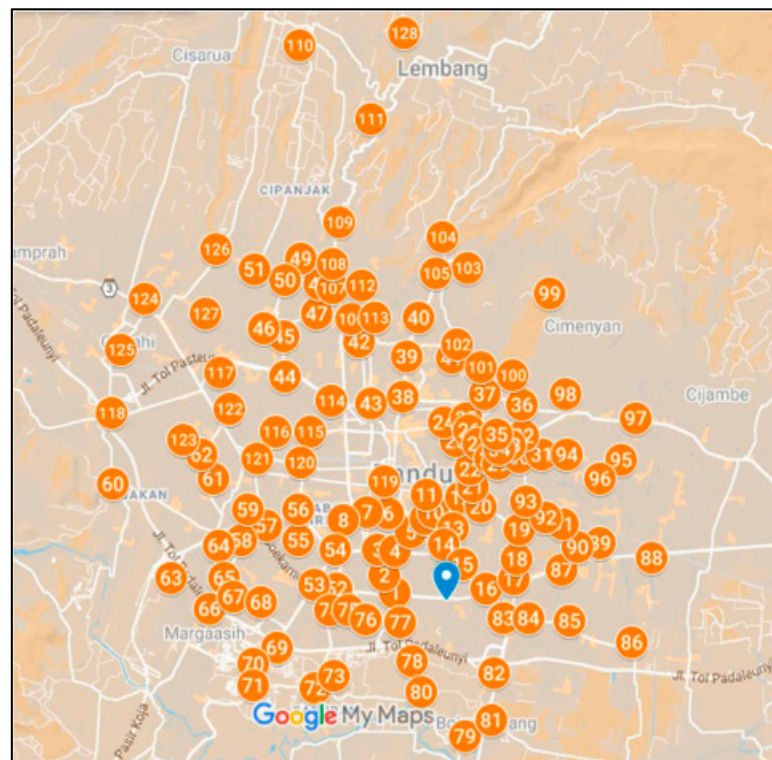
The smaller the MAPE score, the smaller the error of the prediction results, and the greater the MAPE score, the greater the error of the prediction results. The results of a predictive model have an excellent predictive ability if the MAPE score is below 10% and they have a good predictive ability if the MAPE score is between 10% and 20%. Table 1 shows the meaning of the error percentage score in MAPE, where the MAPE score can still be used if it does not exceed 50%; if the MAPE score is above 50%, then the prediction model cannot be used. The MAPE formula is:

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \left| \frac{x_{i,t} - \hat{x}_i}{x_{i,t}} \right| \times 100\%, \quad (10)$$

where $\hat{x}_i = n \frac{\alpha_i}{\alpha_0}$, $\alpha_0 = \sum_{i=1}^m \alpha_i$, $\frac{\alpha_i}{\sum_{i=1}^m \alpha_i} = \theta_i$, and $\sum_{i=1}^m \theta_i = 1, \theta_i > 0$.

3.3. Network Clustering

In the item delivery strategy, the purpose of network clustering is to determine delivery zones (Yudhanegara et al. 2021a, 2021b, 2022). In graph theory, network clustering is a graph partition that divides a graph into more than one subgraph (Elsner 1997; Newman and Girvan 2004). The clustering used in this paper is a recursive spectral bisection (Yudhanegara et al. 2021a, 2022). The network is constructed from the map of Bandung City-Indonesia, in Figure 1, see Yudhanegara et al. (2021a, 2022).

**Figure 1.** Maps of Bandung city as an item delivery network.

In this study, the map is represented as a dynamic network. Dynamic networks experience changes in structure and weight from time to time (Casteigts et al. 2011). The dynamic network concept in this study is defined as a network structure with a change in weight, not a change in the network structure. This dynamic network condition is obtained based on the observation process in alternating periods accompanied by changes in weight. The weight of the network is a random vector with a specific distribution. Unlike the static network, the weights have no distribution.

In Figure 1, the number in the orange circles is the area or location of the destination for the delivery of items, and the blue symbol is the depot. Based on Figure 1, we have the network in Figure 2.

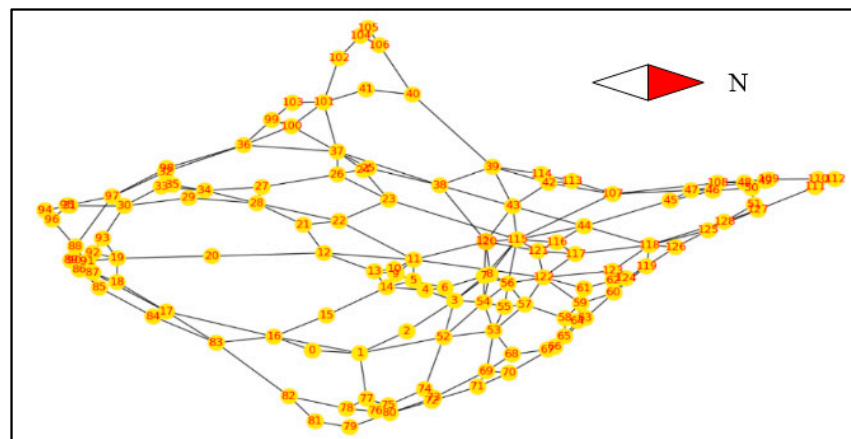


Figure 2. Item delivery network from the map.

In Figure 2, the number in the orange circles is the area or location (node) of the destination for the delivery of items, except for the orange circle with zeros; it's a depot (starting node). Furthermore, Figure 3 is an example of a network clustering into four clusters using a recursive spectral bisection method used in this study (Yudhanegara et al. 2021a).

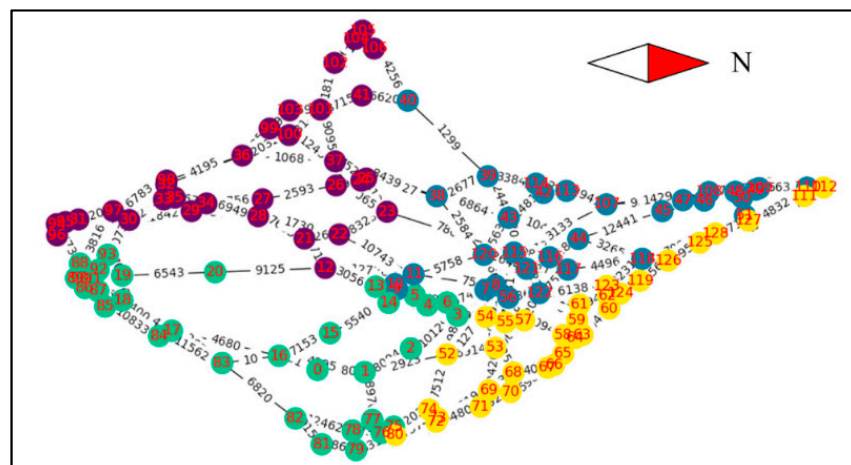


Figure 3. Network clustering with four zones/clusters.

The network in Figure 3 consists of four zones, namely the green zone, gray zone, yellow zone, and purple zone. Green nodes are locations in the green zone, gray nodes in the gray zone, yellow nodes in the yellow zone, and purple nodes in the purple zone.

3.4. Optimization

In this study, the optimization procedure in item delivery uses the Vehicle Routing Problem model, which has been relaxed. This relaxation allows more than one vehicle to pass each road segment. The network is divided into several zones and consists of only one depot. The definition of a delivery zone is a subnet resulting from the clustering of the spectral bisection method. As described previously, the objective function of time optimization in this research problem is used to analyze the structure of the item delivery problem.

Referring to [Dror and Trudeau \(1990\)](#), [Reinelt \(2012\)](#), and [Bernardino and Paias \(2018\)](#), the flow problems with the smallest delivery time used in this study are as follows:

1. One unit of vehicle delivers items to several locations (nodes), but each node is visited only once;
2. The vehicle must return to the depot;
3. The goal is to find the path with the shortest delivery time.

They are adjusted to traffic conditions with changing road density and delivery time limits. The focus of this item delivery process is to get the minimum total time for each zone. The existing data is in the form of the number of vehicles as edge weights. Therefore, it is necessary to convert the edge weights. The conversion is a change from the number of vehicle features to each road segment's travel time feature. The time weight between nodes is obtained by paying attention to traffic parameters, namely speed and traffic density, which refers to the *Direktorat Jenderal Bina Marga* or Department of Highways ([Yudhanegara et al. 2022](#)).

The variables used in the objective function are:

$$\eta_{i,j}^{(q)} : \begin{cases} 1, & \text{if vehicle } q \text{ to node } v_j \text{ from node } v_i \\ 0, & \text{elsewhere.} \end{cases}$$

Next, we have variable $t_{i,j}$, it is the delivery time from v_i to v_j . The goal is to minimize the following function:

$$\mathcal{Z} = \sum_{i=0}^n \sum_{j=0}^n \sum_{q=1}^h t_{i,j} \eta_{i,j}^{(q)}, \quad (11)$$

subject to

$$\sum_{i=0}^n \sum_{q=1}^h \eta_{i,j}^{(q)} = 1, j = 1, 2, \dots, n, i \neq j, \quad (12)$$

and

$$\sum_{j=0}^n \sum_{q=1}^h \eta_{j,k}^{(q)} = 1, k = 1, 2, \dots, n, j \neq k, \quad (13)$$

Equation (12) guarantees that the vehicle that will deliver items to a node originates from the previous node. Equation (13) guarantees that a vehicle that has delivered items from one node must deliver items to the next node. Equations (12) and (13) also indicate that each node can only be stopped precisely once by one vehicle. The other constraints used are:

$$\sum_{j=1}^n \eta_{0,j}^{(q)} = 1, q = 1, 2, \dots, h, \quad (14)$$

and

$$\sum_{j=1}^n \eta_{j,0}^{(q)} = 1, q = 1, 2, \dots, h, \quad (15)$$

where 0 is the depot. Equation (14) indicates that every vehicle's travel route begins at the depot, while Equation (15) indicates that every vehicle's travel route ends at the depot.

Let $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$ be a vector set consisting of n items with $\mathbf{y}_j = [y_{j,1} \ y_{j,2} \ \dots \ y_{j,a}]^T$, $j = 1, 2, \dots, n$ and a are features of the item. Suppose we take a where $y_{j,a} = \tau_j$, which is the delivery time feature. When operating, each vehicle has a different total delivery time, namely T . Suppose that vehicle q has a delivery time capacity of $T^{(q)}$. Each object j to be sent to node $v_j, j = 1, 2, \dots, n$, has the delivery time of $\tau_j \leq \max_q \{T^{(q)}\}$, and each edge $(v_i, v_j), i \neq j$; there is a weight of delivery time between nodes, namely $t_{i,j}$. Thus, the time constraint for model \mathcal{Z} is:

$$\sum_{j=1}^n \tau_j^{(q)} \leq T^{(q)}, \quad q = 1, 2, \dots, h, \tag{16}$$

where

$$\tau_j^{(q)} = \sum_{i=0}^n \sum_{j=0}^n t_{i,j}, \tag{17}$$

and $\tau_j^{(q)}$ is the delivery time of object j to node v_j by vehicle q . Equation (17) guarantees that the delivery time for each vehicle may not exceed the delivery time capacity.

3.5. Item Delivery Strategy

Network clustering aims to divide the network into several delivery zones. The distribution of delivery zones will simplify the delivery process to produce a minimum delivery time according to the available vehicles. In theory, this process avoids deliveries between points that are too far apart, which ensures the route of connected vehicles in the same cluster, and obtains an effective delivery route.

In practice, the item delivery process is influenced by traffic conditions. This condition consists of various factors, one of which is changes in road density, which tend to change or are dynamic, making it difficult to control or predict. These dynamic conditions can be anticipated through a predictive distribution approach. This approach involves statistical inference that predicts delivery zones, as presented in Figure 4.

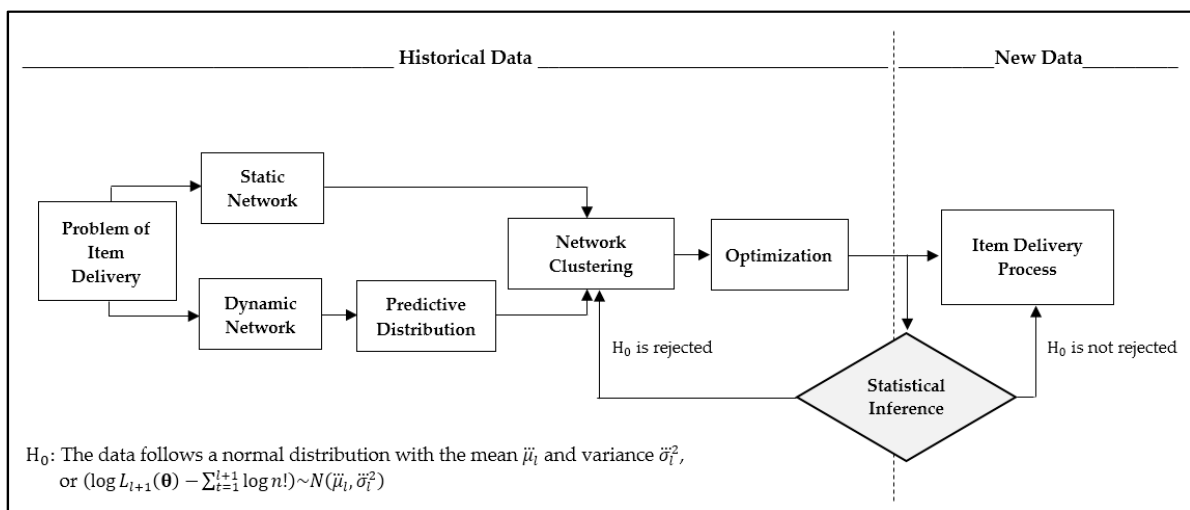


Figure 4. Item delivery strategy.

Figure 4 shows that predictive distribution plays a role in predicting delivery zones, while statistical inference is used to test new data distribution.

4. Result and Discussion

4.1. New Method

To get the distribution of the logarithm of the multinomial likelihood function, we can use Theorem 2.

Theorem 2. (Central Limited Theorem) (Athreya and Lahiri 2006). Let X_1, X_2, \dots, X_m be i.i.d. random samples with mean μ and variance $\sigma^2 > 0$. When m is large, the sample mean $\bar{X}_m = \frac{\sum_{t=1}^m X_t}{m}$ can be approximated by the normal distribution $N\left(\mu, \frac{\sigma^2}{m}\right)$.

Based on Theorem 2, for the multinomial case, we have Theorem 3. It is the distribution of the logarithm of the multinomial likelihood function.

Theorem 3. Let $L_l(\theta) = \prod_{t=1}^l \left(\frac{n!}{\prod_{i=1}^m x_{i,t}!} \prod_{i=1}^m \theta_i^{x_{i,t}}\right)$ be a multinomial likelihood function, for $n, m, l \gg 1$, then random variable $\log L_l(\theta)$ has a normal distribution with mean:

$$\ddot{\mu} = \sum_{t=1}^l \sum_{i=1}^m x_{i,t} \log \theta_i - \sum_{t=1}^l \sum_{i=1}^m \log x_{i,t}! + \sum_{t=1}^l \log n! \tag{18}$$

and variance:

$$\ddot{\sigma}^2 = \frac{m}{(m-1)} \sum_{t=1}^l \sum_{i=1}^m \left(x_{i,t} \log \theta_i - \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m}\right)^2 - \sum_{t=1}^l \sum_{i=1}^m \left(\log x_{i,t}! - \frac{\sum_{i=1}^m \log x_{i,t}!}{m}\right)^2. \tag{19}$$

Proof of Theorem 3. We can find:

$$\begin{aligned} \log L_l(\theta) &= \log \left(\prod_{t=1}^l \left(\frac{n!}{\prod_{i=1}^m x_{i,t}!} \prod_{i=1}^m \theta_i^{x_{i,t}}\right)\right) \\ &= \sum_{t=1}^l \log \left(\frac{n!}{\prod_{i=1}^m x_{i,t}!} \prod_{i=1}^m \theta_i^{x_{i,t}}\right) \\ &= \sum_{t=1}^l \left(\log n! - \sum_{i=1}^m \log x_{i,t}! + \sum_{i=1}^m x_{i,t} \log \theta_i\right) \\ &= \sum_{t=1}^l \left(\log n! - \sum_{i=1}^m \log x_{i,t}! + \sum_{i=1}^m x_{i,t} \log \theta_i\right) \\ &= \sum_{t=1}^l \left(\log n! - m \frac{\sum_{i=1}^m \log x_{i,t}!}{m} + m \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m}\right). \end{aligned}$$

Let $H_t = \frac{\sum_{i=1}^m \log x_{i,t}!}{m} = \overline{\log X_{i,t}!}$ and $O_t = \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m} = \overline{X_{i,t} \log \theta_i}$ be random variables that represent the mean, called the sample mean. Since the sampling is assumed to be a random variable, it has a distribution. The sample mean distribution is the distribution of the means obtained from all possible samples of a population, where the sample size is the same.

The center limit theorem applies to the average sampling distribution. When sampling with a simple random sample size m from a population that comes from any distribution, the distribution of the sample mean can be approximated by a normal distribution provided that the sample size is large—thus, based on the central limit theorem, $H_t \sim N\left(\mu_{H_t}, \sigma_{H_t}^2\right)$, where the mean $\mu_{H_t} \approx \bar{h} = \frac{\sum_{i=1}^m \log x_{i,t}!}{m}$, and variance $\sigma_{H_t}^2 \approx \frac{s^2}{m} = \frac{\sum_{i=1}^m (h_i - \bar{h})^2}{(m^2 - m)}$, $h_i = \log x_{i,t}!$. Thus, we obtained $mH_t \sim N\left(m\mu_{H_t}, m^2\sigma_{H_t}^2\right)$. Next, based on the center limit theorem,

$O_t \sim N(\mu_{O_t}, \sigma_{O_t}^2)$, where the mean $\mu_{O_t} \approx \bar{o} = \frac{\sum_{i=1}^m X_{i,t} \log \theta_i}{m}$, and variance $\sigma_{O_t}^2 \approx \frac{s^2}{m} = \frac{\sum_{i=1}^m (o_i - \bar{o})^2}{(m^2 - m)}$, $o_i = x_{i,t} \log \theta_i$. Thus, we have $mO_t \sim N(m\mu_{O_t}, m^2\sigma_{O_t}^2)$.

Let X_1, X_2, \dots, X_l be i.i.d. random samples, where $X_t \sim N(\mu_{X_t}, \sigma_{X_t}^2)$, $t = 1, 2, \dots, l$. We can find distribution $\sum_{t=1}^l X_t$ through moment generator function (Hogg et al. 2005; Walpole et al. 2012), so $\sum_{t=1}^l X_t \sim N(\sum_{t=1}^l \mu_{X_t}, \sum_{t=1}^l \sigma_{X_t}^2)$. Then, we get $(mH_t + mO_t) \sim N(m\mu_{H_t} + m\mu_{O_t}, m^2\sigma_{H_t}^2 + m^2\sigma_{O_t}^2)$, so we have:

$$(-mH_t + mO_t) \sim N(m(\mu_{O_t} - \mu_{H_t}), m^2(\sigma_{O_t}^2 - \sigma_{H_t}^2)). \tag{20}$$

Next, for $(\log n! - mH_t + mO_t) \sim N(m(\mu_{O_t} - \mu_{H_t}) + \log n!, m^2(\sigma_{O_t}^2 - \sigma_{H_t}^2))$, we find:

$$\sum_{t=1}^l (\log n! - mH_t + mO_t) \sim N(\ddot{\mu}, \ddot{\sigma}^2), \tag{21}$$

where

$$\begin{aligned} \ddot{\mu} &= \sum_{t=1}^l (m(\mu_{O_t} - \mu_{H_t}) + \log n!) \\ &= \sum_{t=1}^l \left(m \left(\frac{\sum_{i=1}^m X_{i,t} \log \theta_i}{m} - \frac{\sum_{i=1}^m \log x_{i,t}!}{m} \right) + \log n! \right) \\ &= \sum_{t=1}^l \sum_{i=1}^m X_{i,t} \log \theta_i - \sum_{t=1}^l \sum_{i=1}^m \log x_{i,t}! + \sum_{t=1}^l \log n!, \end{aligned}$$

and

$$\begin{aligned} \ddot{\sigma}^2 &= \sum_{t=1}^l m^2 (\sigma_{O_t}^2 - \sigma_{H_t}^2) \\ &= \sum_{t=1}^l m^2 \left(\frac{\sum_{i=1}^m (o_i - \bar{o})^2}{l(m-1)} - \frac{\sum_{i=1}^m (h_i - \bar{h})^2}{l(m-1)} \right) \\ &= \sum_{t=1}^l m^2 \left(\frac{\sum_{i=1}^m \left(x_{i,t} \log \theta_i - \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m} \right)^2}{(m^2 - m)} - \frac{\sum_{i=1}^m \left(\log x_{i,t}! - \frac{\sum_{i=1}^m \log x_{i,t}!}{m} \right)^2}{(m^2 - m)} \right) \\ \ddot{\sigma}^2 &= \frac{m}{(m-1)} \left(\sum_{t=1}^l \sum_{i=1}^m \left(x_{i,t} \log \theta_i - \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m} \right)^2 - \sum_{t=1}^l \sum_{i=1}^m \left(\log x_{i,t}! - \frac{\sum_{i=1}^m \log x_{i,t}!}{m} \right)^2 \right). \end{aligned}$$

□

To find the value of $\log n!$, we can approximate the Stirling formula in Equation (22):

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n. \tag{22}$$

We can find Corollary 1 with a more equation from Theorem 3 with the same distribution.

Corollary 1. Let $L_l(\theta) = \prod_{t=1}^l \left(\frac{n!}{\prod_{i=1}^m x_{i,t}!} \prod_{i=1}^m \theta_i^{x_{i,t}} \right)$ be a multinomial likelihood function, for $n, m, l \gg 1$, then random variable $(\log L_l(\theta) - \sum_{t=1}^l \log n!)$ has a normal distribution with mean:

$$\ddot{\mu} = \sum_{t=1}^l \sum_{i=1}^m x_{i,t} \log \theta_i - \sum_{t=1}^l \sum_{i=1}^m \log x_{i,t}!, \tag{23}$$

and variance $\ddot{\sigma}^2 = \ddot{\sigma}^2$, Equation (19).

Proof of Corollary 1. From $\log L_l(\theta)$ we have:

$$\log L_l(\theta) = \sum_{t=1}^l \left(\log n! - \sum_{i=1}^m \log x_{i,t}! + \sum_{i=1}^m x_{i,t} \log \theta_i \right),$$

and thus,

$$\log L_l(\theta) - \sum_{t=1}^l \log n! = \sum_{t=1}^l \left(- \sum_{i=1}^m \log x_{i,t}! + \sum_{i=1}^m x_{i,t} \log \theta_i \right).$$

Next, based on Equations (19) and (20), we have:

$$(-mH_t + mO_t) = \sum_{t=1}^l \left(- \sum_{i=1}^m \log x_{i,t}! + \sum_{i=1}^m x_{i,t} \log \theta_i \right).$$

Thus, we get:

$$\sum_{t=1}^l \left(- \sum_{i=1}^m \log x_{i,t}! + \sum_{i=1}^m x_{i,t} \log \theta_i \right) \sim \left(m(\mu_{O_t} - \mu_{H_t}), m^2(\sigma_{O_t}^2 - \sigma_{H_t}^2) \right),$$

suppose

$$\ddot{\mu} = m(\mu_{O_t} - \mu_{H_t}) = \sum_{t=1}^l \sum_{i=1}^m x_{i,t} \log \theta_i - \sum_{t=1}^l \sum_{i=1}^m \log x_{i,t}!$$

and

$$\begin{aligned} \ddot{\sigma}^2 &= m^2(\sigma_{O_t}^2 - \sigma_{H_t}^2) \\ &= \frac{m}{(m-1)} \left(\sum_{t=1}^l \sum_{i=1}^m \left(x_{i,t} \log \theta_i - \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m} \right)^2 \right. \\ &\quad \left. - \sum_{t=1}^l \sum_{i=1}^m \left(\log x_{i,t}! - \frac{\sum_{i=1}^m \log x_{i,t}!}{m} \right)^2 \right). \end{aligned}$$

□

As an illustration, the values of $\log L_l(\theta)$ where $m = 38, n = 2500, l = 30$, for $\ell = 2000$ times, are shown in Figure 5.

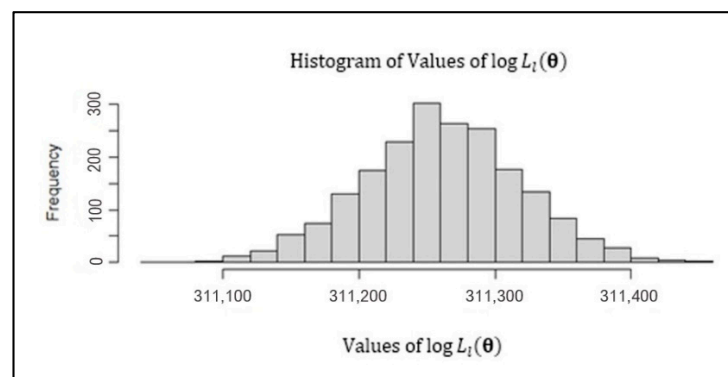


Figure 5. Histogram of values of $\log L_l(\theta)$ for $\ell = 2000$.

Next, as an illustration, the values of $\log L_l(\theta)$, where $m = 260, n = 1,738,665, l = 30$, for $\ell = 1000$, are shown in Figure 6.

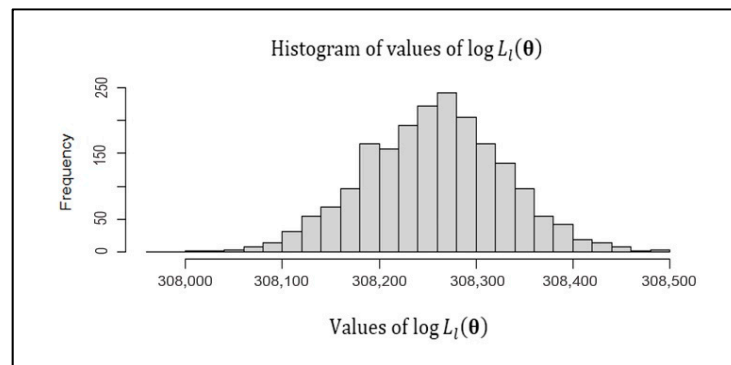


Figure 6. Histogram of values of $\log L_l(\theta)$ for $l = 1000$.

The probability density function for histogram in Figure 5 is:

$$f(x) = \frac{1}{9435\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-311,261}{9435} \right)^2}, \tag{24}$$

and

$$f(x) = \frac{1}{4505\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-308,256}{4505} \right)^2} \tag{25}$$

for the histogram in Figure 6. Based on Figures 5 and 6, the histogram of values of $\log L_l(\theta)$ are symmetrical. They are theoretically corresponding to Theorem 3. Corollary 1 state that $(\log L_l(\theta) - \sum_{t=1}^l \log n!) \sim N(\ddot{\mu}, \ddot{\sigma}^2)$. It is the statistical form for multinomial goodness-of-fit test. It applies the following hypothesis test:

H₀: $(\log L_{l+1}(\theta) - \sum_{t=1}^{l+1} \log n!)$ follows a normal distribution with the mean $\ddot{\mu}$ and variance $\ddot{\sigma}^2$,

H₁: $(\log L_{l+1}(\theta) - \sum_{t=1}^{l+1} \log n!)$ does not follow a normal distribution with the mean $\ddot{\mu}$ and variance $\ddot{\sigma}^2$.

If we look at the logarithmic form of the likelihood function for $l + 1$ observations, then Corollary 2 applies. This result illustrates that there is no need to store historical data in online testing, but it is enough to use the last calculation.

Corollary 2. Value $\log L_{l+1}(\theta) - \sum_{t=1}^{l+1} \log n!$, $\ddot{\mu}_{l+1}$ and $\ddot{\sigma}_{l+1}^2$, it can be obtained iteratively through $\log L_l(\theta) - \sum_{t=1}^l \log n!$, $\ddot{\mu}_l$ and $\ddot{\sigma}_l^2$; thus:

$$\log L_{l+1}(\theta) - \sum_{t=1}^{l+1} \log n! = \left(\log L_l(\theta) - \sum_{t=1}^l \log n! \right) + \left(\sum_{i=1}^m \log x_{i,(l+1)}! + \sum_{i=1}^m x_{i,(l+1)} \log \theta_i \right),$$

$$\ddot{\mu}_{l+1} = \ddot{\mu}_l + \left(\sum_{i=1}^m x_{i,(l+1)} \log \theta_i - \sum_{i=1}^m \log x_{i,(l+1)}! \right),$$

and

$$\ddot{\sigma}_{l+1}^2 = \ddot{\sigma}_l^2 + \frac{m}{(m-1)} \left(\sum_{i=1}^m \left(x_{i,(l+1)} \log \theta_i - \frac{\sum_{i=1}^m x_{i,(l+1)} \log \theta_i}{m} \right)^2 \right) - \sum_{i=1}^m \left(\log x_{i,(l+1)}! - \frac{\sum_{i=1}^m \log x_{i,(l+1)}!}{m} \right)^2.$$

Proof of Corollary 2. Consider the logarithm of the multinomial likelihood function for l observations as follows:

$$\begin{aligned} \log L_l(\theta) &= \sum_{t=1}^l \left(\log n! - m \frac{\sum_{i=1}^m \log x_{i,t}!}{m} + m \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m} \right) \\ &= \log L_{l-1}(\theta) + \left(\log n! - \sum_{i=1}^m \log x_{i,l}! + \sum_{i=1}^m x_{i,l} \log \theta_i \right). \end{aligned}$$

Iteratively, for $l + 1$ observations, based on the above form, the logarithm of the likelihood function can be written as:

$$\log L_{l+1}(\theta) = \log L_l(\theta) + \left(\log n! - \sum_{i=1}^m \log x_{i,(l+1)}! + \sum_{i=1}^m x_{i,(l+1)} \log \theta_i \right).$$

Thus, based on the equation, we get:

$$\log L_{l+1}(\theta) - \sum_{t=1}^{l+1} \log n! = \left(\log L_l(\theta) - \sum_{t=1}^{l+1} \log n! \right) + \left(\sum_{i=1}^m \log x_{i,(l+1)}! + \sum_{i=1}^m x_{i,(l+1)} \log \theta_i \right).$$

The iterative process also applies to $\ddot{\mu}_{l+1}$ and $\ddot{\sigma}_{l+1}^2$, and is described as follows:

$$\begin{aligned} \ddot{\mu}_l &= \sum_{t=1}^l \sum_{i=1}^m x_{i,t} \log \theta_i - \sum_{t=1}^l \sum_{i=1}^m \log x_{i,t}! \\ &= \sum_{t=1}^{l-1} \sum_{i=1}^m x_{i,t} \log \theta_i + \sum_{i=1}^m x_{i,l} \log \theta_i - \sum_{t=1}^{l-1} \sum_{i=1}^m \log x_{i,t}! - \sum_{i=1}^m \log x_{i,l}! \\ &= \sum_{t=1}^{l-1} \sum_{i=1}^m x_{i,t} \log \theta_i - \sum_{t=1}^{l-1} \sum_{i=1}^m \log x_{i,t}! + \left(\sum_{i=1}^m x_{i,l} \log \theta_i - \sum_{i=1}^m \log x_{i,l}! \right) \\ &= \ddot{\mu}_{l-1} + \left(\sum_{i=1}^m x_{i,l} \log \theta_i - \sum_{i=1}^m \log x_{i,l}! \right). \end{aligned}$$

Based on the description above, we have:

$$\ddot{\mu}_{l+1} = \ddot{\mu}_l + \left(\sum_{i=1}^m x_{i,(l+1)} \log \theta_i - \sum_{i=1}^m \log x_{i,(l+1)}! \right).$$

For variance, we have:

$$\begin{aligned} \ddot{\sigma}_l^2 &= \frac{m}{(m-1)} \left(\sum_{t=1}^l \sum_{i=1}^m \left(x_{i,t} \log \theta_i - \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m} \right)^2 - \sum_{t=1}^l \sum_{i=1}^m \left(\log x_{i,t}! - \frac{\sum_{i=1}^m \log x_{i,t}!}{m} \right)^2 \right) \\ &= \frac{m}{(m-1)} \sum_{t=1}^l \sum_{i=1}^m \left(x_{i,t} \log \theta_i - \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m} \right)^2 - \sum_{t=1}^l \sum_{i=1}^m \left(\log x_{i,t}! - \frac{\sum_{i=1}^m \log x_{i,t}!}{m} \right)^2 \\ &= \frac{m}{(m-1)} \left(\sum_{t=1}^{l-1} \sum_{i=1}^m \left(x_{i,t} \log \theta_i - \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m} \right)^2 + \sum_{i=1}^m \left(x_{i,l} \log \theta_i - \frac{\sum_{i=1}^m x_{i,l} \log \theta_i}{m} \right)^2 \right. \\ &\quad \left. - \left(\sum_{t=1}^{l-1} \sum_{i=1}^m \left(\log x_{i,t}! - \frac{\sum_{i=1}^m \log x_{i,t}!}{m} \right)^2 + \sum_{i=1}^m \left(\log x_{i,l}! - \frac{\sum_{i=1}^m \log x_{i,l}!}{m} \right)^2 \right) \right) \\ &= \frac{m}{(m-1)} \left(\sum_{t=1}^{l-1} \sum_{i=1}^m \left(x_{i,t} \log \theta_i - \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m} \right)^2 - \sum_{t=1}^{l-1} \sum_{i=1}^m \left(\log x_{i,t}! - \frac{\sum_{i=1}^m \log x_{i,t}!}{m} \right)^2 \right. \\ &\quad \left. + \sum_{i=1}^m \left(x_{i,l} \log \theta_i - \frac{\sum_{i=1}^m x_{i,l} \log \theta_i}{m} \right)^2 - \sum_{i=1}^m \left(\log x_{i,l}! - \frac{\sum_{i=1}^m \log x_{i,l}!}{m} \right)^2 \right) \\ \ddot{\sigma}_l^2 &= (\ddot{\sigma}_{l-1}^2) + \frac{m}{(m-1)} \left(\sum_{i=1}^m \left(x_{i,l} \log \theta_i - \frac{\sum_{i=1}^m x_{i,l} \log \theta_i}{m} \right)^2 - \sum_{i=1}^m \left(\log x_{i,l}! - \frac{\sum_{i=1}^m \log x_{i,l}!}{m} \right)^2 \right). \end{aligned}$$

Thus, we find:

$$\ddot{\sigma}_{l+1}^2 = \ddot{\sigma}_l^2 + \frac{m}{(m-1)} \left(\sum_{i=1}^m \left(x_{i,(l+1)} \log \theta_i - \frac{\sum_{i=1}^m x_{i,(l+1)} \log \theta_i}{m} \right)^2 - \sum_{i=1}^m \left(\log x_{i,(l+1)}! - \frac{\sum_{i=1}^m \log x_{i,(l+1)}!}{m} \right)^2 \right).$$

□

The hypothesis test used is the Z statistical test with the formula in Equation (26):

$$z_{statistical} = \frac{\left(\log L_{l+1}(\theta) - \sum_{t=1}^{l+1} \log n! \right) - \ddot{\mu}}{\ddot{\sigma}}, \tag{26}$$

where $\log L_{l+1}(\theta) - \sum_{t=1}^{l+1} \log n!$ is random variable under the assumption H_0 . The decision-making criteria H_0 is not rejected when $-z_{\frac{\alpha}{2}} \leq z_{statistical} \leq z_{\frac{\alpha}{2}}$, H_0 is rejected if $z_{statistical} < -z_{\frac{\alpha}{2}}$ or $z_{\frac{\alpha}{2}} < z_{statistical}$. The margin of error for the confidence interval $(1 - \alpha)$ in statistics on Equation (26) can be determined through Theorem 4. This theorem provides a formula for the error tolerance limit for the population average at the confidence interval $(1 - \alpha)$.

Theorem 4. Let $(\log L_{l+1}(\theta) - \sum_{t=1}^{l+1} \log n!) \sim N(\ddot{\mu}_l, \ddot{\sigma}_l^2)$ be a random variable, where $\ddot{\mu}_l = \sum_{t=1}^l \sum_{i=1}^m x_{i,t} \log \theta_i - \sum_{t=1}^l \sum_{i=1}^m \log x_{i,t}!$ and $\ddot{\sigma}_l^2 = \frac{m}{(m-1)} \sum_{t=1}^l \sum_{i=1}^m \left(x_{i,t} \log \theta_i - \frac{\sum_{i=1}^m x_{i,t} \log \theta_i}{m} \right)^2 - \sum_{t=1}^l \sum_{i=1}^m \left(\log x_{i,t}! - \frac{\sum_{i=1}^m \log x_{i,t}!}{m} \right)^2$, then margin of error (ε) for the confidence interval $(1 - \alpha)$ is $\varepsilon = z_{(1-\frac{\alpha}{2})} \ddot{\sigma}_l$.

Proof of Theorem 4. Let ε be a margin of error for the confidence interval $(1 - \alpha)$. Consider the following form:

$$\begin{aligned} P\left(\left| \left(\log L_{l+1}(\theta) - \sum_{t=1}^{l+1} \log n! \right) - \ddot{\mu}_l \right| < \varepsilon \right) &= 1 - \alpha \\ P\left(-\varepsilon < \left(\left(\log L_{l+1}(\theta) - \sum_{t=1}^{l+1} \log n! \right) - \ddot{\mu}_l \right) < \varepsilon \right) &= 1 - \alpha \\ P\left(-\frac{\varepsilon}{\ddot{\sigma}_l} < \frac{\left(\left(\log L_{l+1}(\theta) - \sum_{t=1}^{l+1} \log n! \right) - \ddot{\mu}_l \right)}{\ddot{\sigma}_l} < \frac{\varepsilon}{\ddot{\sigma}_l} \right) &= 1 - \alpha \\ P\left(-\frac{\varepsilon}{\ddot{\sigma}_l} < z_{stat} < \frac{\varepsilon}{\ddot{\sigma}_l} \right) &= 1 - \alpha. \end{aligned}$$

Based on the description, we have:

$$\begin{aligned} \frac{\varepsilon}{\ddot{\sigma}_l} &= z_{(1-\frac{\alpha}{2})} \\ \varepsilon &= z_{(1-\frac{\alpha}{2})} \ddot{\sigma}_l. \end{aligned}$$

□

The steps for testing new data based on historical data are described in Algorithm 1.

Algorithm 1. Data testing.

1. Input $(\ddot{\mu}, \ddot{\sigma}^2)$
2. Find the estimated parameter (θ) of the predictive distribution (based on historical data).
3. Enter the new observed value under the assumption H_0 .
4. Perform the Z statistical test at the significance level α . H_0 is not rejected, it means that $(l + 1) \sim N(\ddot{\mu}, \ddot{\sigma}^2)$.
5. Update the parameters (θ) and $(\ddot{\mu}, \ddot{\sigma}^2)$ if H_0 is rejected.
6. Return to step 3.

4.2. Limitations

The smallest value or minimum limit of m, n , and l is determined based on the simulation of the values of $\log L(\theta)$ so that Theorem 3 and Corollary 1 can be used. In the next step, the values of $\log L(\theta)$ are tested for model fit using chi-squared statistics under the assumption H_0 , H_0 : data follow a normal distribution. The rejection criterion H_0 is based on a p -value, i.e., the probability value of obtaining a result is at least as extreme as the observed result of a statistical hypothesis test, assuming that the null hypothesis is true. The p -value is used as an alternative rejection point to provide a minor significance level with the null hypothesis being rejected. Obtaining a smaller p -value indicates more substantial evidence supporting the alternative hypothesis.

In the first simulation, for the value of $m = 9, l = 10$, and $n = 115$, for $\ell = 500$, the histogram is obtained in Figure 7a. Based on the model fit test, a p -value < 0.0001 is obtained so that at a significance level of 0.05, the value $\log L(\theta)$ does not follows a normal distribution. Then, for the values of $m = 10, l = 8$, and $n = 115$, for $\ell = 500$, the histogram is obtained in Figure 7b. A p -value = 0.018 was obtained based on the model fit test. At the 0.05 significance level, it can be stated that the $\log L(\theta)$ value does not follow a normal distribution, similar to the previous simulation in Figure 7a.

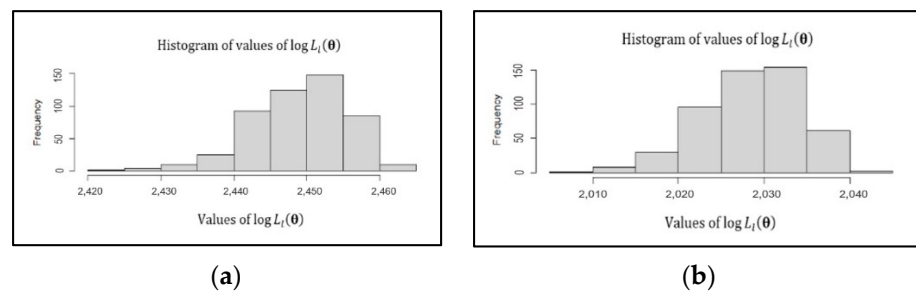


Figure 7. Histogram of values of $\log L_l(\theta)$, (a) for $\ell = 500, m = 9, l = 10, n = 115$; and (b) for $\ell = 500, m = 10, l = 8, n = 115$.

The following simulation is carried out for the values of $m = 10, l = 9$, and $n = 115$ with $\ell = 500$; the histogram is obtained in Figure 8a.

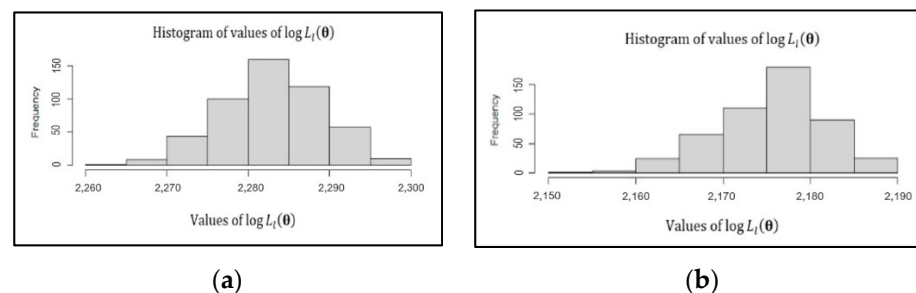


Figure 8. Histogram of values of $\log L_l(\theta)$, (a) for $\ell = 500, m = 10, l = 9, n = 115$; and (b) for $\ell = 500, m = 10, l = 9, n = 10$.

From these values, based on the model fit test, the p -value = 0.066. This means that at a significance level of 0.05, it can be concluded that the value of $\log L(\theta)$ follows a normal distribution with mean $\hat{\mu} = 2283.0220$ and variance $\hat{\sigma}^2 = 11.877$.

Next, a simulation is carried out for the values of $m = 10$, $l = 9$, and $n = 110$, for $k = 500$, to obtain the histogram in Figure 8b. Based on the model fit test, a p -value < 0.0001 was obtained with these values. Thus, at a significance level of 0.05, the value of $\log L(\theta)$ does not follow a normal distribution. After that, the last simulation was carried out for the values of $m = 10$, $l = 9$, and $n = 111$, for $k = 500$; the histogram is shown in Figure 9.

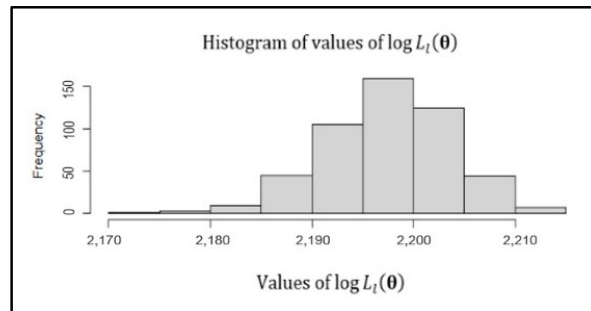


Figure 9. Histogram of values of $\log L_l(\theta)$, for $k = 500$, $m = 10$, $l = 9$, $n = 111$.

Based on the model fit test, the p -value = 0.363 at a significance level of 0.05 means that the value of $\log L(\theta)$ follows a normal distribution with probability density function:

$$f(x) = \frac{1}{1.969\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-2197.444}{1.969}\right)^2}. \tag{27}$$

This simulation concludes that in the case of multinomial distribution, Theorem 3 and Corollary 1 apply to the minimum values of $m = 10$, $l = 9$, and $n = 111$.

4.3. Simulation

Let $\mathbf{X}_t = [X_{1,t} \ X_{2,t} \ \dots \ X_{m,t}]^T$ be a random vector that stating the number of vehicles that might pass each road at t th day, where $X_{i,t}$ is a random variable stating the number of vehicles pass i th road at t th day. Then, assume \mathbf{X}_t follows a multinomial distribution with the parameter θ (probability). Furthermore, a total of 1,738,665 events are obtained with 260 event categories.

The network is constructed from the map of Bandung City-Indonesia, shown in Figure 1. We applied them to transportation data, namely traffic density. In this case, a map is presented in a dynamic network, and categories are represented as roads (Yudhanegara et al. 2020, 2021a, 2021b). The total number of events is the total number of vehicles passing on the road, see Figure 3 (Yudhanegara et al. 2021a).

Then, the data are generated 40 times under the assumption of a multinomial distribution with the probability mass function of $\mathbf{X}_t \sim \text{Mult}(\theta_1, \theta_2, \dots, \theta_{260}; 1,738,665)$, namely:

$$p(\mathbf{x}; \theta) = \frac{1,738,665!}{\prod_{i=1}^{260} x_{i,t}!} \prod_{i=1}^{260} \theta_i^{x_{i,t}}, \tag{28}$$

where $\sum_{i=1}^{260} \theta_i = 1$ and $\sum_{i=1}^{260} x_{i,t} = 1,738,665$. The parameter θ is generated randomly from the standard uniform distribution ($U(0,1)$). We have 40 generated data, where 60% (24 datas) was used as training data, and 40% (16 datas) was used as testing data. The expected value and the estimated parameter of training data are obtained from the predictive distribution in Equation (6).

The training data that has been used as the network weight is presented in Figure 10. Figure 10 shows the network that has been clustered into four zones using the spectral

bisection method. Each zone tends to change cluster members within its zone at different times. Furthermore, the data prediction simulation is presented in Table 2.

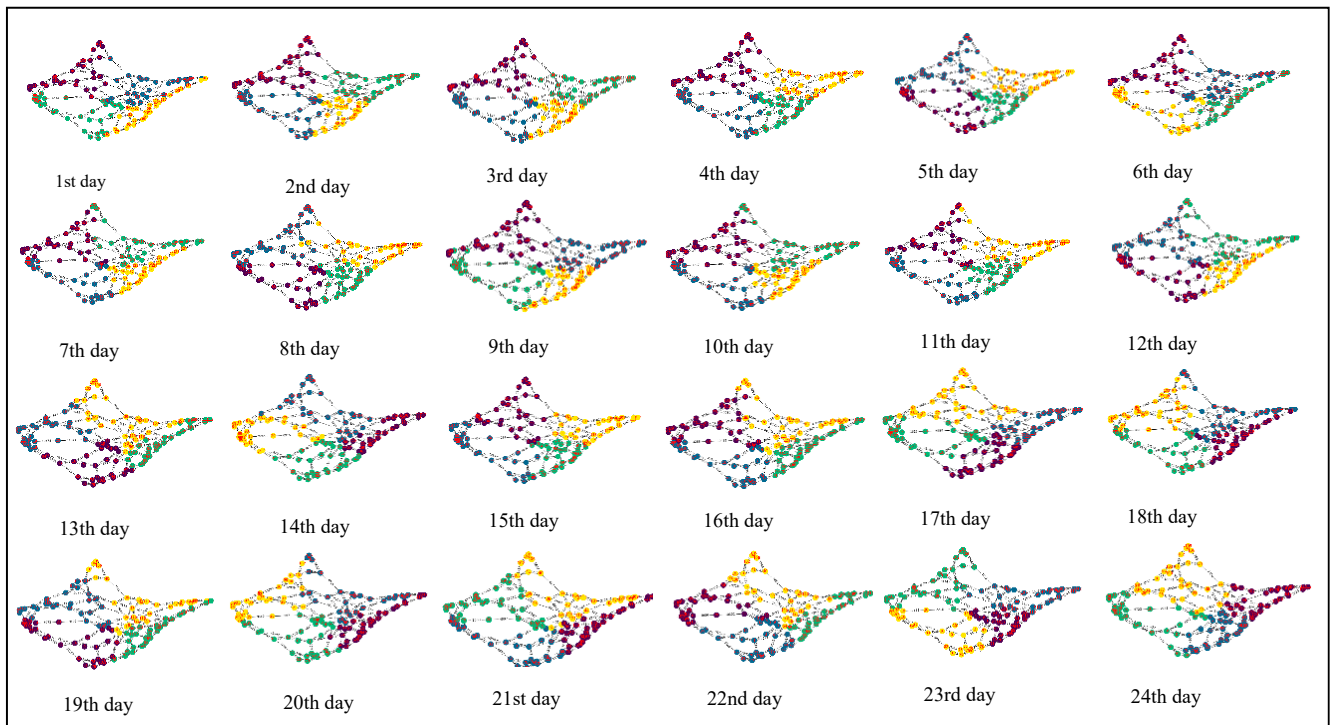


Figure 10. Network clustering at different times.

By using the estimated parameter (θ) of historical data, the value of $Z_{statistical}$ is obtained, as shown in Table 2.

Table 2. Summary of hypothesis testing with $\alpha = 0.05$, and $z_{critical} = 1.96$ or $z_{critical} = -1.96$.

Parameter (θ) and ($\ddot{\mu}, \ddot{\sigma}^2$)	Data	$z_{statistical}$	Decision	Item Delivery Strategy Changes	Parameter ($\ddot{\mu}_t, \ddot{\sigma}_t^2$)	Error (ϵ)
D_{1-24}	D_{1-24} vs. D_{25}	-0.154	H_0 is not rejected	No	D_{1-24}	0.23912
D_{1-24}	D_{1-24} vs. D_{26}	-0.308	H_0 is not rejected	No	D_{1-25}	0.22540
D_{1-24}	D_{1-24} vs. D_{27}	-0.462	H_0 is not rejected	No	D_{1-26}	0.21168
D_{1-24}	D_{1-24} vs. D_{28}	-0.615	H_0 is not rejected	No	D_{1-27}	0.23324
D_{1-24}	D_{1-24} vs. D_{29}	-0.769	H_0 is not rejected	No	D_{1-28}	0.21168
D_{1-24}	D_{1-24} vs. D_{30}	-0.923	H_0 is not rejected	No	D_{1-29}	0.24108
D_{1-24}	D_{1-24} vs. D_{31}	-1.003	H_0 is not rejected	No	D_{1-30}	0.25872
D_{1-24}	D_{1-24} vs. D_{32}	-1.263	H_0 is not rejected	No	D_{1-31}	0.24892
D_{1-24}	D_{1-24} vs. D_{33}	-1.465	H_0 is not rejected	No	D_{1-32}	0.23912
D_{1-24}	D_{1-24} vs. D_{34}	-1.671	H_0 is not rejected	No	D_{1-33}	0.23324
D_{1-24}	D_{1-24} vs. D_{35}	-1.865	H_0 is not rejected	No	D_{1-34}	0.22932
D_{1-24}	D_{1-24} vs. D_{36}	-2.005	H_0 is rejected	Yes	D_{1-35}	0.31752
D_{1-36}	D_{1-36} vs. D_{37}	-0.007	H_0 is not rejected, by updating (θ) and ($\ddot{\mu}, \ddot{\sigma}^2$)	No	D_{1-36}	0.19404
D_{1-36}	D_{1-36} vs. D_{38}	-0.271	H_0 is not rejected	No	D_{1-37}	0.2156
D_{1-36}	D_{1-36} vs. D_{39}	-0.429	H_0 is not rejected	No	D_{1-38}	0.2410
D_{1-36}	D_{1-36} vs. D_{40}	-0.533	H_0 is not rejected	No	D_{1-39}	0.2001

Note: D_{25} is the 25th data, and D_{1-24} is the 1st to 24th data.

Next, to evaluate the prediction results of the predictive distribution model, we use MAPE. The results of the MAPE scores calculation is described in Table 3.

Table 3. The results of the MAPE calculation and interpretation for the prediction.

Data	MAPE Score	Interpretation	Data	MAPE Score	Interpretation
E _{1–24} vs. D ₂₅	0.9%	very accurate	E _{1–24} vs. D ₃₃	9.8%	very accurate
E _{1–24} vs. D ₂₆	1.4%	very accurate	E _{1–24} vs. D ₃₄	8.8%	very accurate
E _{1–24} vs. D ₂₇	1.0%	very accurate	E _{1–24} vs. D ₃₅	7.8%	very accurate
E _{1–24} vs. D ₂₈	1.0%	very accurate	E _{1–36} vs. D ₃₇	4.3%	very accurate
E _{1–24} vs. D ₂₉	1.9%	very accurate	E _{1–36} vs. D ₃₈	3.2%	very accurate
E _{1–24} vs. D ₃₀	2.8%	very accurate	E _{1–36} vs. D ₃₉	2.2%	very accurate
E _{1–24} vs. D ₃₁	4.3%	very accurate	E _{1–36} vs. D ₄₀	2.8%	very accurate
E _{1–24} vs. D ₃₂	5.7%	very accurate			

Note: D₂₅ is the 25th data, and E_{1–24} is the expected value of the 1st to 24th data.

In general, based on the criteria for the MAPE scores in Table 1, the MAPE scores between the observed value and the expected value in Table 3 indicates that the model used is accurate because the MAPE scores are below 10%.

4.4. Optimization as Risk Management

The next goal of risk management is to provide information about potential sources of risk in the company (Lam 2016). Risk management is an effort to reduce risk in technical implementation and business decision making (Lam 2016). The purpose of risk management is to mitigate or track sources that have the potential to threaten business productivity and security. This tracking process can be carried out by research and procedural analysis of every company activity, from production to asset management. When risks are found and analyzed, it is necessary to make efforts so that risks do not occur and threaten business continuity.

Although it has extended and successive stages, the risk management process is one of the essential components of business management that can protect the company from many problems. In this study, particularly the item delivery strategy, optimization is one of the most critical components of risk management (Cooke and Pinter 1989; Better et al. 2008).

Next, we analyze the optimization of a dynamic network. Based on Table 2, for H₀ is rejected, if there is no change in the item delivery strategy (ignoring H₀ is rejected), then the optimization of delivery time in each zone will no longer be effective (Yudhanegara et al. 2022). This happens because the routes used are not correct. The result is significant delivery times on certain vehicles, while others are too small. The optimization of the time in the order of cases for H₀ is rejected as described in Tables 4 and 5.

Table 4. The risks when ignoring H₀ is rejected (for the yellow zone and gray zone).

Yellow Zone						Gray Zone					
Total time on initial strategy			Total time when ignoring H ₀ is rejected			Total time on initial strategy			Total time when ignoring H ₀ is rejected		
Vehicle 1	Vehicle 2	Vehicle 3	Vehicle 1	Vehicle 2	Vehicle 3	Vehicle 1	Vehicle 2	Vehicle 3	Vehicle 1	Vehicle 2	Vehicle 3
86.42 min	45.878 min	58.751 min	120.42 min	170.3 min	70.5 min	73.448 min	136.21 min	152.18 min	37.6 min	50.9 min	40.41 min

Table 5. The risks when ignoring H₀ is rejected (for the green zone and purple zone).

Green Zone						Purple Zone					
Total time on initial strategy			Total time when ignoring H ₀ is rejected			Total time on initial strategy			Total time when ignoring H ₀ is rejected		
Vehicle 1	Vehicle 2	Vehicle 3	Vehicle 1	Vehicle 2	Vehicle 3	Vehicle 1	Vehicle 2	Vehicle 3	Vehicle 1	Vehicle 2	Vehicle 3
56.518 min	53.063 min	58.6235 min	154.6 min	40.9 min	40.87 min	70.03 min	46.302 min	50.13 min	98.1 min	140.3 min	145.7 min

5. Conclusions

Let $\mathbf{X}_t = [X_{1,t} X_{2,t} \dots X_{m,t}]^T$ be a random vector, $\mathbf{X}_t \sim \text{Mult}(\theta_1, \theta_2, \dots, \theta_m, n)$, $t = 1, 2, \dots, l$, for $n, m, l \gg 1$, for the appropriate model fit test is used, namely the Z statistical

test with $\log L_l(\theta) \sim N(\ddot{\mu}, \ddot{\sigma}^2)$ or $(\log L_l(\theta) - \sum_{i=1}^l \log n!) \sim N(\ddot{\mu}, \ddot{\sigma}^2)$. The algorithm presented as the method of inferential statistics for online decisions. It is suitable for streaming data, some of which consist of traffic density data. This method can predict data with small error values, resulting in very accurate predictions. The method does not require large memory storage for historical data.

The case presented in this paper is suitable for controlling traffic jam in cities and traffic jam on holidays. The solution to the problems presented in this paper can be helpful for the transportation department to make decisions or provide information to the public. The impact on the item delivery company by having accurate information on traffic jams will enable them to formulate strategies for item deliveries.

For future research, we can simulate the data based on the predictive distribution of the multivariate Poisson cases. Moreover, a case study can be used with other discrete multivariate distributions. For example, there are cases where n is not constant, and $n \rightarrow \infty$, the probability of the occurrence of the event being considered is $\theta \rightarrow 0$, and $n \cdot p = \omega$, where ω is the rate, so that further research is related to the multivariate Poisson distribution. Then, we can determine a new test method built based on the logarithm of the likelihood function for the multivariate Poisson distribution.

After obtaining the predictive distribution for the multivariate Poisson case, we will also apply it to education. Clustering for category data and decision-making methods will be needed for analyzing the risk of failing to graduate from university. Examples of data can be seen in a study conducted by [Yudhanegara and Lestari \(2019\)](#).

Author Contributions: Conceptualization, S.W.I. and M.R.Y.; methodology, S.W.I.; software, M.R.Y.; validation, S.W.I., K.N.S., and M.R.Y.; formal analysis, S.W.I.; investigation, K.N.S.; resources, M.R.Y.; data curation, K.N.S.; writing—original draft preparation, M.R.Y.; writing—review and editing, S.W.I.; visualization, K.N.S.; supervision, S.W.I.; project administration, S.W.I.; funding acquisition, S.W.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Riset Unggulan LPPM ITB 2022 (Second Year), grant number NIP 197508041999031003 and the APC was funded by Sapto Wahyu Indratno.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are generated under the assumption of a multinomial distribution with our program. They are available on request from the corresponding author.

Acknowledgments: The authors would like to thank LPPM ITB for research funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Aranburu, Larraitz, Laureano F. Escudero, M. Araceli Garín, María Merino, and Gloria Pérez. 2016. Risk Management for Mathematical Optimization under Uncertainty. *Documento de Trabajo BILTOKI 1*: 1–18.
- Athreya, Krishna B., and Soumendra N. Lahiri. 2006. *Measure Theory and Probability Theory*. New York: Springer.
- Avetisyan, Marianna, and Jean-Paul Fox. 2012. The Dirichlet-multinomial model for multivariate randomized response data and small samples. *Psicologica* 33: 362–90.
- Azen, Razia, and Cindy M. Walker. 2010. *Categorical Data Analysis for the Behavioral and Social Sciences*. New York: Routledge Taylor & Francis Group.
- Bernardino, Raquel, and Ana Paia. 2018. Solving the family traveling salesman problem. *European Journal of Operational Research* 267: 453–66. [\[CrossRef\]](#)
- Better, Marco, Fred Glover, Gary Kochenberger, and Haibo Wang. 2008. Simulation optimization: Applications in risk management. *International Journal of Information Technology & Decision Making* 7: 571–87. [\[CrossRef\]](#)
- Brock, William A., and Steven N. Durlauf. 2003. Multinomial Choice with Social Interactions. *Technical Working Paper: National Bureau of Economic Research* 288: 1–44.
- Casteigts, Arnaud, Paola Flocchini, Walter Quattrociocchi, and Nicola Santoro. 2011. Time-varying graphs and dynamic networks. *International Conference on Ad-Hoc Networks and Wireless* 27: 346–59.
- Cooke, Roger, and Janos Pinter. 1989. Optimization in risk management. *Civil Engineering Systems* 6: 122–28. [\[CrossRef\]](#)

- Degtereva, Viktoria, Maria Liubarskaia, Viktoria Merkusheva, and Alexey Artemiev. 2022. Increasing Importance of Risk Management in the Context of Solid Waste Sphere Reforming in Russian Regions. *Risks* 10: 79. [CrossRef]
- Dror, Moshe, and Pierre Trudeau. 1990. Split delivery routing. *Naval Research Logistic* 37: 383–402. [CrossRef]
- Elsner, Ulrich. 1997. *Graph Partitioning*. Chemnitz: Technische Universität Chemnitz.
- Hogg, Robert V., Joseph W. McKean, and Allen T. Craig. 2005. *Introduction to Mathematical Statistics*, 6th ed. Hoboken: Pearson.
- Johnson, Norman L., Samuel Kotz, and Narayanaswamy Balakrishnan. 1996. *Discrete Multivariate Distributions*. New York: John Wiley & Sons Inc.
- Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 2012. *Loss Models: From Data to Decisions*, 4th ed. Hoboken: John Wiley & Sons Inc.
- Krokhmal, Pavlo, Michael Zabarankin, and Stan Uryasev. 2011. Modeling and optimization of risk. *Surveys in Operations Research and Management Science* 16: 49–66. [CrossRef]
- Lam, James. 2016. *Strategic Risk Management: Optimizing the Risk Return Profile*. New York: The Association of Accounts and Financial Professionals in Business (IMA).
- Lewis, Colin D. 1982. *Industrial and Business Forecasting Methods*. London: Butterworths.
- Nelissen, Heleen E., Daniella Brals, Hafsat A. Ameen, Marijn van Der List, Berber Krammer, Tanimola M. Akande, Wendy Janssens, and Anja H. Van't Hoog. 2020. The prominent role of informal medicine vendors despite health insurance: A weekly diaries study in rural Nigeria. *Health Policy and Planning* 35: 54–63. [CrossRef] [PubMed]
- Newman, Mark E. J., and Michelle Girvan. 2004. Finding and evaluating community structure in networks. *Physical Review E* 69: 026113. [CrossRef] [PubMed]
- Nownes, Anthony J. 1992. Primaries, general elections, and voter turnout: A multinomial logit model of the decision to vote. *American Politics Quarterly* 20: 205–26. [CrossRef]
- Oosterhoff, John, and Willem R. Van Zwet. 1972. The likelihood ratio test for the multinomial distribution. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* 1: 31–49. [CrossRef]
- Owen, Art B. 2018. *Introduction to Statistical Inference: Generalized Likelihood Ratios Test*. Stanford: Stanford University, Available online: <https://web.stanford.edu/class/archive/stats/stats200/stats200.1172/Lecture22.pdf> (accessed on 23 March 2020).
- Reddy, Byreddy R., Ch. Deepika, G. Navya, and Gopal Raja. 2015. Factors effecting the usage of public transport multinomial logit model. *International Journal of Computer Science and Information Technologies* 6: 4672–675.
- Reinelt, Gerhard. 2012. TSPLIB: A traveling salesman problem library. *ORSA Journal on Computing* 3: 376–84. [CrossRef]
- Rith, Monorom, Fengqi Liu, Pai-Hsien Hung, Kento Yoh, Alexis M. Fillone, and Jose B. M. Biona. 2019. R programming language written for estimation of the integrated multinomial logit-linear regression model based on a copula approach: A technical article. *Proceedings of the Eastern Asia Society for Transportation Studies* 12: 1–11.
- Swamidass, Paul M. 2000. *Encyclopedias of Production and Manufacturing Management*. Boston: Kluwer Academic Publishing.
- Walpole, Ronald E., Raymond H. Myers, Sharon L. Myers, and Keying Ye. 2012. *Probability and Statistics for Engineer and Scientists*, 6th ed. Hoboken: Pearson.
- Yudhanegara, Mokhammad R., and Karunia E. Lestari. 2019. Clustering for multi-dimensional data set: A case study on educational. *Journal of Physics: Conference Series* 1280: 042025. [CrossRef]
- Yudhanegara, Mokhammad R., Sapto W. Indratno, and R. R. Kurnia N. Sari. 2020. Clustering for items distribution network. *Journal of Physics: Conference Series* 1496: 012019. [CrossRef]
- Yudhanegara, Mokhammad R., Sapto W. Indratno, and R. R. Kurnia N. Sari. 2021a. Dynamic items delivery network: Prediction and clustering. *Heliyon* 7: e06934. [CrossRef]
- Yudhanegara, Mokhammad R., Sapto W. Indratno, and R. R. Kurnia N. Sari. 2021b. Role of clustering method in items delivery optimization. *Journal of Physics: Conference Series* 2084: 012011. [CrossRef]
- Yudhanegara, Mokhammad R., Sapto W. Indratno, and R. R. Kurnia N. Sari. 2022. Prediction of traffic density and item delivery strategy through clustering. *Sesiomadika* 5: 27–46.