Link-data-based approximation of path travel time distribution with Gaussian copula estimated through Lasso

Ke Wan: PhD candidate, Phone: 609-356-8466, Fax: 609-258-1563
Email: kwan@princeton.edu, corresponding author;
Department of Operations Research and Financial Engineering
Princeton University, Princeton, NJ 08544 USA

Alain Kornhauser: Professor, Phone: 609-258-4657
Fax: 609-258-1563, Email: alaink@princeton.edu
Department of Operations Research and Financial Engineering
Princeton University, Princeton, NJ 08544 USA

Word count: 7000
Figure count: 10
Table count: 12
July 31, 2009
Abstract

In this paper, we highlight the characteristics of floating car data, and extend our research on single link to pathes.

By the law of total probability, we proposes to approximate the total path travel time distribution by the probability-weighted sum of a series of conditional path travel time distributions, conditioning on the sequence of entering time to the links of the path which is expressed as its corresponding (entering-time) lag vector. Any of such conditional distribution is the sum distribution of a sequence of link travel time distribution with specific dependence structure. The dependent structure is modeled by a lagged Gaussian copula while the marginal distributions are estimated by kernel method. The L1-constrain-minimization(Lasso) method is utilized to obtain an invertible covariance matrix of the Gaussian copula even for limited data.

Compare to the iterative type procedure, this approach is efficient when the number of scenarios to visit is limited and it resolve the "conditional distribution puzzle" in the iterative formulas.

Keyword: Path Travel Time Distribution, Sum of Dependent Random Variables, Gaussian Copula, Lasso, Lagged Copula, Lag Vector
1 Introduction

Rigorous and efficient computational procedures to the shortest path problem, one of the fundamental problems of Operations Research, are evolving to be at the core of everyday life. Today, millions simply tell a system, be it an in-car Personal Navigation Device or a cell phone, his or her destination. The system "instantly" provides detailed turn-by-turn directions by solving a shortest path problem. Invariably, these systems make the heroic assumption that "costs" throughout the network are non-changing and deterministic. This enormously simplifying assumption makes the problem tractable and "solvable" but doesn't address the real problem: What are the turn-by-turn directions that will allow one to travel to one's destination in the best way given that costs throughout the network are neither constant nor deterministic. This is especially true for daily commuters who "know" "all" of the various ways to get to and from work but unfortunately have no means of knowing or anticipating what today's traffic conditions as they are distributed throughout the various alternatives will play out to a "best" way to go. Elements of this "real" stochastic optimization problem are the focus of this paper. What makes this "real" problem difficult are the following:

1. the cost (say, travel time) on individual arcs are not scalars but instead time varying probability distributions
2. the time varying probability distributions on different arcs are not independent as is often assumed.
3. the cumulative cost of any path through the network is characterized by a probability distribution, thus any comparison of paths requires the comparison of probability distributions, not simple scalar values

A common way to address the interdependence of costs on neighboring arcs is to study the Pearson's linear correlation which assumes that a joint normal distribution (elliptical family) is form of the dependent structure between the costs on different arcs.

Difficulties place severe burdens on the information systems required to generate the solution procedure's data needs and the computational procedures required to accommodate the additional reality. To begin to address these difficulties we propose the following:

1. We consider the joint distribution of travel time in the usual non-normal case: the marginal distribution is not necessarily normal and the dependence structure
between them either. Then path travel time is approximated by the sum of dependent travel times which the traveler experiences at a series of specific entering time to the links in the path.

2. Parameter estimation can be conducted through Lasso method, which tackle the ill-conditioned estimation when data is limited.

3. In our approach, we continue to use decision rules/risk measures based on the travel time distribution, such as mean-variance rule, stochastic dominance rules etc. By comparison of these decision statistics, we believe, better objective decisions can be offered for travelers, and routing choices can be further optimized.

4. For real time applications, the estimator need to be applied as the travel goes on in a trip and can be applied for the arc replety over time. The approximation simplified down the procedure and enable iterative decision making.

5. The amount of OD pairs in the actual road network is large, and between each pair there are numerous paths. Estimation will be based on links instead of path. Denote $N$ as the number of links in a large network, the data storage should be of order $O(N)$.

Based on the work in Wan(2009)[23]. This paper will mainly focus on aggregating the travel time distribution/utility of links to get that of the paths and possible decision rule selection.

The paper is organized as follows: In Section 2, A literature review on related theory is given, including copula theory and estimation of covariance matrix. In Section 3, Our Framework and models are given: dependence structure estimation, and interpretations of decision rules are presented. In Section 4, the aggregation of the link travel time is designed and compared. In Section 5, Conclusion and future work is discussed.

2 Methodology Clarification and Related Theory

In this section, we review different literature and offer the theoretical base for a solution to the fundamental problem.

2.1 Path travel time estimation

Travel time estimation has been a hot topic in transportation research, and numerous studies have focused on the accurate prediction of travel time of highways. Including:
density estimator Petty(1998)[16], cross-validation estimator, Zhang(2003)[24], time series in Vemuri(1998) [21], local linear models in Dailey(2000)[5] and CTM, Daganzo(1994)[4], Delay function in Nie(2005)[14]. Most of research is about link travel time estimation and the research about the distribution estimation is limited. Wan(2008)[23] proposes a copula-based link travel time distribution estimator, it reconstruct the joint distribution of the travel time between different links, and yields proper estimation.

The further question is how to aggregate link travel time distribution to get path travel time distribution, some more literature is reviewed below:

In Chen(1968)[3], link-based model and path-based model are compared. And path based model is preferred. However there is not relation between link and path. The path based model is totally based on the observed total path travel time. For real ATIS system it is impossible to store the total path travel time for each path because the issue of combinatorial issue.

The simplest assumption is to assume travel times on all the links along a path are generated by distributions that are statistically independent. And various modification is made to the results made by this assumption.

Raha(2006)[17] uses coefficient of variation as

$$CV = \frac{\mu}{\sigma}$$

, to derive the estimates for the path travel time variance. There are five methods for the estimation of path travel-time variance from its component segment travel-time variances. And the assumption is statistically independence between arcs and the estimates are:

$$\sigma_p^2 = \sum_{l \in L(p)} \sigma_l^2$$

$$\sigma_p^2 = \frac{(\sum_{l \in L(p)} \mu_j)^2}{m^2} \sum_{l \in L(p)} \frac{\mu_j}{\sigma_j}$$

$$\sigma_p^2 = \left( \sum_{l \in L(p)} \mu_j \right)^2 med_{l \in L(p)} \frac{\mu_j}{\sigma_j}$$

$$\sigma_p^2 = \left( \sum_{l \in L(p)} \mu_j \right)^2 \frac{1}{2} \left( \max_{l \in L(p)} \frac{\mu_j}{\sigma_j} + \min_{l \in L(p)} \frac{\mu_j}{\sigma_j} \right)$$

The last three estimators are derived considering the trip CV as as the mean/median/mean of max and min of the CV over all segment realizations in the path.

Sherali(2006)[18] derived another formulation for estimating the path travel-time
variance. They used the maximum and minimum segment travel-time CVs to construct bounds on the path CV given that the CV is independent of the length of the segment. And the
\[
\sigma_p^2 = \frac{(\sum \mu_j)^2}{\sum \mu_j^2} \sum_{l \in L(p)} \sigma_l^2
\]
It is a modification from the above estimation based on the mean travel time.

In Fu(1998)[9], the assumption on the link travel time is that the travel times on individual links at a particular point in time are statistically independent. He argued the probability distributions of the link travel times are modeled as functions of the time of day, the time of day correlation between individual links is explicitly taken into account. However this way can not capture the dynamic dependent structure of travel time.

In Waller(2002)[22] the assumption is one step further the one-step dependence between arcs

In the paper Pattan(2003)[15], a Gaussian kernel is used to estimate the continuous mean travel time at a particular point in time t, a locally three-point polynomial approximation is used to estimate the mean link travel time as a function of time of day and a two factor model, in which stochastic travel time is caused by a systematic error and a vehicle error is used for error decomposition. when considering the case where travel times on two consecutive links are dynamic and stochastic. To estimate the conditional mean and variance of one on the other, the joint probability density function is required. "However, from a practical standpoint it is impossible to estimate this function." Then it use some approximation to deal with it. In our research, we focus on the estimation of such joint distribution function.

Among the usual assumption on the dependent structure, joint normal distribution is most usual one with respect to correlated arc travel time estimation. Dailey(2000)[5], Gao(2006)[10] considers the simplified correlation model over time but not on different arcs and still does not go too far. An improved technic to tackle the non-normal joint distribution is in need.

2.1.1 Dependence structure and Copula theory

Dependence structure is the dependent relationship between different random variables, mathematically, it is expressed in copula functions Nelsen(2006)[13], Cheruini(2004)[20].The above research is based on certain assumptions on dependence structure so it is a special case of the dependence structure theory in this section.
Definition 1. Defining $V_H(B) = H(x_2, y_2) - H(x_1, y_2) - H(x_2, y_1) + H(x_1, y_1)$ for $B = [x_1, y_1] \times [x_2, y_2]$, a 2-place real function $H$ is 2-increasing if $V_H(B) \geq 0$ for all rectangles $B$ whose vertices lie in $\text{Dom } H$.

Definition 2. Let $A_i$ be subsets of $I$, denote $a_i$ the least element of $A_i$, The function $H$ is grounded if for every $(v, z)$ of $A_1 \times A_2$,

$$H(a_1, z) = 0 = H(v, a_2)$$

Definition 3. A two-dimensional sub-copula is a function $C'$ with the following properties:

1. $\text{Dom } C' = S_1 \times S_2$, where $S_1$ and $S_2$ are subsets of $I = [0, 1]$, containing 0 and 1;
2. $C'$ is grounded and 2-increasing
3. For every $u$ in $S_1$ and every $v$ in $S_2$, $C'(u, 1) = u$ and $C'(1, v) = v$.

Definition 4. A two-dimensional copula is a two-dimensional sub-copula with $S_1 = S_2 = I$

The definitions above can be extended to multivariate case and therefore a multivariate copula is a multivariate distribution function defined on the unit cube $[0, 1]^n$, with uniformly distributed margins.

A copula-based approach allows a decomposition of a joint distribution into its marginal distributions and its copula. On the other hand marginal distributions may be combined to a joint distribution assuming a specific copula. The crucial point in using a copula-based approach is that it allows for a separate modeling of the marginal distributions (i.e. the univariate travel time distributions) and the dependence structure (the copula).

Sklar(1959) shows that an n-dimensional joint distribution function may be decomposed into its n marginal distributions and a copula, which completely describes the dependence between the n variables. The fundamental mathematical theorem here is as follows: Nelsen(2006) [13]

Theorem 1. sklar’s thm of copula

Let $H$ be a joint distribution function with margins $F$ and $G$. Then there exists a copula $C$ such that for all $x, y$ in $\mathbb{R}$,

$$H(x, y) = C(F(x), G(y))$$
if \( F \) and \( G \) are continuous, then \( C \) is unique; otherwise, \( C \) is uniquely determined on \( \text{Ran} F \times \text{Ran} G \). Conversely, if \( C \) is a copula and \( F \) and \( G \) are distribution functions, then the function \( H \) defined by is joint distribution function with margins \( F \) and \( G \).

For the conditional copula, the definitions is given by the following:

**Definition 5.** A conditional bivariate distribution function is a right continuous function \( H : \mathbb{R}^2 \rightarrow [0, 1] \) with the properties:

1. \( H(x, -\infty|F) = H(-\infty, y|F) = 0 \) and \( H(\infty, \infty|F) = 1 \)
2. \( V_H([x_1, x_2] \times [y_1, y_2]) = H(x_2, y_2|F) - H(x_1, y_2|F) - H(x_2, y_1|F) + H(x_1, y_1|F) \geq 0 \)

for all \( x_1, x_2, y_1, y_2 \in \mathbb{R} \) and \( x_1 \leq x_2, y_1 \leq y_2 \).

Where \( F \) is the conditioning sigma-algebra.

And for the conditional copula we have the following theorem:

**Theorem 2.** Sklar’s theorem of conditional copula:

Sklar’s Theorem for Continuous Conditional Distributions

Let \( H \) be a conditional bivariate distribution function with continuous margins \( F \) and \( G \), and let \( F \) be some conditioning set. Then there exists a unique conditional copula \( C : [0, 1] \times [0, 1] \) such that

\[
H(x, y|F) = C(F(x|F), g(y|F)|F), \forall x, y \in \mathbb{R}
\]

Conversely, if \( C \) is a conditional copula and \( F \) and \( G \) are the conditional distribution functions of the two random variables \( X \) and \( Y \), then the function \( H \) defined by the equation is a bivariate conditional distribution function with margins \( F \) and \( G \).

Note, different dependency measures such as person correlation, spearman’s \( \rho \) and kendall’s \( \tau \) can be unified under the copula theory, see Nelson(1995) [12] and Nelson Fredricks(2006) [7]. Copula theory enables us to go beyond the linear Pearson correlation coefficient and the joint Gaussian approximation, as for the practical traffic data, joint Gaussian is not always the case.

### 2.1.2 Covariance matrix estimation and the Lasso method

The problem when estimates the covariance matrix or the gaussian matrix \( P \), is how to deal with the limited non-synchronized data with different amounts on different link pairs.

1. The marginal link travel time data is abundant.
2. There is some data for pair wise correlation between links.
3. There is limited synchronized data for all the links under consideration. The data matrix $X^tX$ is nearly ill-conditioned.

The first way based on relatively sufficient pairwise data is to estimate the pair-wise correlation and modify the term into 0 if it is not significant. The classical significant test for correlation is

$$t = \frac{\rho}{\sqrt{(1 - \rho^2)/(N - 2)}}$$

distributed approximately as T distribution with df=N-2. When the test shows the correlation is insignificant we can set it to 0, to make the dependence structure simple. This method can take full use of data, but this term-by-term estimates may lead to non-positive semi-definite covariance matrixes. This brings problem when we proceed to estimate the joint distribution.

The more reliable way is to estimate the covariance matrix directly using the observed travel time for all the links in the path. And the empirical estimator of covariance matrix is

$$\Sigma = \frac{1}{n}X^TX - \frac{1}{n}X^T11^TX$$

The estimator leads to positive definite matrix but it yields nearly ill-conditioned covariance matrix when the number of independent data is approximately equal to the dimension and this can lead to large error in estimation.

As a further improvement, two types of advanced version is introduced, shrinkage estimators and penalized estimator.

1. The shrinkage estimator is in a linear combination of the covariance matrix targeting to a prior.

$$\Sigma_e = \alpha \Sigma_0 + (1 - \alpha)\Sigma$$

where $\Sigma_0$ is a prior positive definite covariance matrix. This estimator will always give good positive definite covariance matrix. And its error is subject to the prior matrix selected and might be large when data limited.

2. Since the floating car data may be limited for many paths, we prefer another way of estimation stated as follows:

In Meinshausen(2006)[11], the problem of inverse covariance matrix estimates can be formulated as an optimization problem with the objective:

$$\max_{\Omega: \Omega \geq 0} \log \det \Omega + tr(S\Omega) + \rho|\Omega|_1$$
where $\$\$ is the sample variance matrix and $\rho$ is the weight. Its dual problem can be formulated into a $L - 1$ constrained optimization problem, called Lasso.

In the paper, it also shows that neighborhood selection with the Lasso is a computationally attractive alternative to standard covariance selection for sparse high-dimensional graphs. In Friedman(2008)[8], lasso is used to estimate the inverse covariance matrix. The penalized term can reduce unnecessary terms in the covariance matrix and hence make the estimate robust and accurate. It is better than term-by-term significant test in the sense of ensuring a positive semi-definite covariance matrix, while needs more simultaneous observations on all the links in the path. This is a tradeoff when we do the estimation for the gaussian-type copulas.

In Banerjee(2008)[1], the author formulates the graph discovering problem in a similar way to get the solution sparse. They propose an algorithm uses block coordinate descent algorithm, which can be interpreted as recursive 1-norm penalized regression(Lasso).

For the lasso algorithm, it is first proposed in Tibshirani[19].It minimizes sum of square s subject to the sum of the absolute value of the coefficient being less than a constant.

$$\begin{align*}
(\bar{\alpha}, \bar{\beta}) &= \operatorname{argmin} \sum_{i=1}^{n} (y_i - \alpha \sum_j \beta_j x_{ij})^2 \\
\text{s.t} \\
\sum_j |\beta_j| &\leq t
\end{align*}$$

And a solution method is given in a iterative procedure which starts from overall least square estimates and solve constrained least square each step. No research has applied lasso to do parameter estimation for gaussian copula, nor applied it in the traffic research area. Our research tries to tackle this problem.

3 Decision based on path travel time estimation

The problem under consideration is the stochastic routing decision problem. In our research, we try to reconstruct the joint distribution of arc travel time using appropriate copula models models and further address the dependence structure with time series model if data is sufficient, and use factor models when data is not sufficient. Finally, several decision rules are given to yield a best routing choice. The settings are given as below:
3.1 Problem modeling and structuring

The basic settings in the research is as follows:

1. We consider the stochastic of links and dependence between links non-normal case, each link has a marginal distribution $F_i$. The links are dependent with certain joint structure, not necessarily joint normal.

2. We only store the data for each link. The data includes the travel time observations and the corresponding time to make them. So the data is stored in links and restricted to O(N) where $N$ is the number of links.

3. For each traveler, we have finite number of pathes to consider. That is the method focus on the routing decision with finite subset of pathes in a indefinite large network. The design aims at aggregating the link travel time distribution to get path travel time distribution.

4. We consider dynamic system which might dynamically change and we make assumptions on periodical changes and constancy in short time intervals, if needed for certain procedures.

5. For decision rules, they are within the framework set up by Wan(2009)[23], decisions are made based on conditional distribution estimated for each user.

Next we continue with the objective function and compare different formulations:

Denote a network as $N, L$ described by a finite set $N$ of nodes and a set $L = \subseteq N \times N$ of links between those nodes. A Path is a sequence of links $p_{ab} = l_1, \ldots , l_n|l_i \in L & l_1 \in L_a & l_n \in L_b$. and the path set for given OD pair $P_{ab} = p|allp_{ab}$. Then the Minimum risk decision problem is:

$$\min_p U(p)$$

s.t

$$U(p) = \int f(p)\mu(dp)$$

where

$f(p)$ is the value of the decision statistics when path travel time equals a certain value. We do not use the expectation $E(U(p))$ here as some of the statistics can not be represented in this form.

$\mu(dp)$ is probability distribution of the path travel time it is in the following form:
We define the random travel time on a link entering at time $t$ as $T_t$

$$\sum_{i=1}^{N} T_{s_i}$$

where

$$s_i = \sum_{j=1}^{i-1} T_{s_j} \text{ and } s_1 = 0$$

To estimate such path travel time, the most intuitive way is to study the sequence of link travel time experienced by any specific user. This way needs the trip-by-trip data, and it will need too much storage.

An alternative way is to estimate the path travel time distribution based on link travel time data, and there are several ways to achieve this. For all methods proposed, the enumeration of possible scenario of link travel time is inevitable, and the problem is how to cut out unnecessary scenario or how to achieve a better approximation at a given number of state enumeration. We analyze the three ways to do the estimation:

Conditioning on the starting time the trip $t_1$, the forward procedure is:

$$P(\sum_{i=1}^{n} T_{t_i} \leq t) = \int P(T_{t_n} = t - s | \sum_{i=1}^{n-1} T_{t_i} \leq s) P(\sum_{i=1}^{n-1} T_{t_i} = s)ds \text{ for each } t$$

with the initial condition: $P(T_{t_2} + T_{t_1} \leq t) = \int P(T_{t_2} = t - s | T_{t_1} \leq s) P(T_{t_1} = s)ds$

when $n = 2$.

Conditioning on the starting time the trip $t_1$, the backward procedure is:

$$P(\sum_{i=n}^{N} T_{t_i} \leq t) = \int P(T_{t_n} = s | \sum_{i=1}^{n-1} T_{t_i} = x) P(\sum_{i=n+1}^{N} T_{t_i} \leq t - s | T_{t_n} = s, \sum_{i=1}^{n-1} T_{t_i} = x)ds \text{ for each } t$$

with the initial condition: $P(\sum_{i=N-1}^{N} T_{t_i} \leq t) = \int P(T_{t_{N-1}} = s | T_{t_N} \leq t - s | T_{t_{N-1}} = s)ds \text{ when } n = N - 1$.

However, to carry out these two schemes, there are two serious drawbacks:

1. Numerical error is large when we carried it out in continuous multidimensional integration. If using a discrete state enumeration, too many scenarios need to be enumerated and estimation error can be overwhelming with limited data.

A simple analysis which assumes the distribution and conditional distribution are all stationary over time will indicate the problem of state enumeration: if for a path made of $M$ links, each link takes value in $N$ discrete intervals, for each conditional distribution distribution of the Path travel time we consider it takes
value in \( N \) intervals. The total states to visit is \( N^M \).

The resulting estimation is still a discrete distribution on \( N \) intervals and contains large error, while this discrete distribution converges to the true distribution as \( N \to \infty \).

2. The "Conditional distribution puzzle": The conditional distribution in the iteration formula is hard to estimate exactly. The assumption about the one step dependence of link travel time Waller[22] is just a usual way to simplify this conditional distribution structure and it brings great errors.

The third way to do approximation of the path travel time based on link data is to estimate the conditional path distribution for each lag vector and use continuous distribution to approximate the conditional path travel time distribution. The definitions and formulas are as follows:

\[
P(T_{t_1} + T_{t_2} + \ldots + T_{t_N} < t) = \int P(T_{t_1} + T_{t_2} + \ldots + T_{t_N} < t|V)\lambda(dV) \quad (3)
\]

where

\( V = (t_1, t_2, \ldots, t_n) \) is the lag vector, denote the sequence of entering time to the links in the path.

\( \lambda(dV) \) is the probability corresponding to a given lag vector \( V \)

**In continuous scheme:**

\[
P(T_{t_1} + T_{t_2} + \ldots + T_{t_N} < t|V) = 1_{(t_1 + t_2 + \ldots + t_N < t)} \quad (4)
\]

is a conditional path travel time given the entering time to the links is a certain lag vector \( V \).

\( 1_{(t_1 + t_2 + \ldots + t_N < t)} \) is the indicator function takes value 1 if \( t_1 + t_2 + \ldots + t_N < t \), 0 otherwise.

**In discrete scheme:**

\[
P(T_{t_1} + T_{t_2} + \ldots + T_{t_N} < t|V) = \int 1_{(T_{t_1} + T_{t_2} + \ldots + T_{t_N} < t)}\mu(V, \Delta, x) \quad (5)
\]

\[
\mu(V, \Delta, x) = P(T_{t_1} + T_{t_2} + \ldots + T_{t_N} < t|V \in P_{V, \Delta}) = C(F_{T_{t_1}}(x_1), F_{T_{t_2}}(x_2), \ldots, F_{T_{t_n}}(x_n)|\Delta) \quad (6)
\]
is the cumulative probability function for the link travel time, conditioning on a given
lag vector $V$ and scheme $\Delta$.

$\Delta = (\Delta_1, \Delta_2, \ldots, \Delta_n)$ is the discrete interval vector, i th element is for the interval
length of Link i.

$P_{V, \Delta} = [t_1 - \Delta_1/2, t_1 + \Delta_1/2] \times [t_2 - \Delta_2/2, t_2 + \Delta_2/2] \times \ldots \times [t_N - \Delta_N/2, t_N + \Delta_N/2]$

is the sub space associate with $V$.

$C$ is a suitable copula between the link travel time distributions $F_i$ for the random
travel time $T_i$.

$F_{T}(T_i(x))$ is the value at $x_i$ of the cumulative distribution function of the conditional
link travel time for $T_i$, corresponding to a given lag vector $V$.

$[F_{T1}(x_1), \ldots, F_{TN}(x_N)]$, should satisfy $E(T_i) = t_{i+1} - t_i$. In other word, each of
them is the conditional link travel time distribution given the mean travel time is
t_{i+1} - t_i. It is just a decomposition of the total distribution into specific scenarios.

After we get the aggregation procedure for fixed lags, the ultimate estimation will
be a linear combination of several basic conditional distributions. The weights is in
proportion to the probability of the occurrence of the corresponding lag vectors, i.e:

$$\sum_{i=1}^{N} \lambda_i P(T_{t_1} + T_{t_2} + \ldots + T_{t_N} < t | V_i, \Delta)$$

where $\lambda_i = \frac{P(V=V_i)}{\sum_{j=1}^{N} P(V=V_j)}$

To clarify the probability assumptions made here, discussion about the choice of
$\Delta_i$ and the lag Vector $V$ is necessary:

1. The fundamental assumptions is that the Consistency of marginal distribution
(CMS) and Constancy of dependence structure(CDS) in Wan(2009)[23] will be
assumed to hold for fixed time intervals $\Delta_i$.

2. The lag vector and $\Delta$ should match the choice of the marginal distribution. We
choose the marginal distribution for link i such that $\mu_i = t_{i+1} - t_i$ and $\sigma = \frac{\Delta T}{m} \ m \in \mathbb{R}^+$

3. The set of all lag vectors in the discrete scheme will be a set such that

$$\Theta = \left\{ V : \Omega = \bigcap P_v \text{ and } P_{v_i} \bigcap P_{v_j} = \Phi \text{ if } i \neq j \right\}$$

where $\Omega = \{ (t_1, \ldots, t_N) \ , \ t_i \in \mathbb{R}^+ \ , \forall i \}$. Conditional distribution conditioning
on $V$ is for the pathes which lie in $P_v$. By definition, when $\delta \rightarrow 0 \ \Theta \rightarrow \Omega$
4. We approximate the conditional path travel time if each link travel time $T_i$ is experienced by entering Link $i$ in the time interval $[t_i - \Delta_i/2, t_i + \Delta_i/2]$. $\Delta_i$ here acts as a tradeoff between discrete scheme and continuous process.

When $\Delta T \to 0$ the conditional distribution converges to a specific sequence of link travel time along a trip, when $\Delta_i \approx \sigma T_i$, this is to assume the probability law of travel time hold constant in the $\Delta_i$ and the procedure capture all the data in the discrete scenario to estimate the underlying relation with a continuous structure. This mixed design enables a good estimation of travel time in limited enumeration. Due to the sparsity of floating car data, this $Delta T$ discrete scheme is efficient to find the data in need.

An intuitive explanation for the definition above is the conditional distribution for a given lag vector will be the distribution of the experienced travel time for a user in a scenario which satisfies the following conditions:

1. The traveler enters the i-th link at $t_i$, and experienced travel time $T_{t_i}$, the conditional expectation of which is $t_{i+1} - t_i$.
2. $T_{t_i}$s are constant in their joint probability law in the interval $\Delta_i$ for the time intervals $[t_i - \Delta T/2, t_i + \Delta T/2]$.
3. The dependent structure between all the $T_{t_i}$ satisfying 1 and 2 is determined by the time lags vector $V$ and tolerance $\Delta$ and is constant as the starting time $t_1$ shifts forwards.

Compare to the iterative scheme, this yields better approximation when the scenarios visited is small. We will show the approximation by just visiting one scenario (The one the link travel time all takes value in a interval near its expectation). This is a great advantage for realtime systems, and the more scenarios are analyzed the better the approximation will be, by the law of total probability.

This design is the second layer of the three-layer structure of routing decision framework based on copula methods, first given in Wan(2009) [23]. It is restated as follows:

1. Travel time distribution estimation for links with limited data and the basic framework to use copula method in the context of floating car data processing. This is shown in Wan(2008) [23].
2. Aggregate the travel time distribution/utility for paths and path decision making. The routing decision is then based on the estimated path travel time. This is the main issue in this paper.
3. Upper level learning scheme. Based on path travel time, a rolling horizon decision scheme is the first step. Optimal routing selection based on learning over time is a further development. This will be given in the succeeding papers.

In the next sections, we will focus on the construction of conditional path distribution w.r.t the following specific fixed lag vector and the lags in the vector are just the overall expected travel time in each link. We define this as the main scenario for the path, and $V$ is

$$t_i = \sum_{j=1}^{i-1} ET_{ij} \text{ and } t_1 = 0$$ (8)

The corresponding cumulative distribution used is the overall cumulative distribution $F_{X_i}$ and the tolerance $\Delta_i$ is set as fixed constants, $C$ times standard deviation of the current link travel time distribution. Notice here, the overall cumulative travel time distribution for Link $i$ satisfies the condition that $E(T_{ti}) = t_{i+1} - t_i$. For link $i$, more specific distributions which satisfies the same condition can be used as a surrogate, such as the conditional distribution on link $i$ at time $t$ whose mean is $t_{i+1} - t_i$.

4 Estimation of conditional path travel time

Follow the definition of discussion of previous sections, the estimation of the path travel time includes three steps:

1. Data organization considering time lag along a path;
2. Joint distribution estimation based on organized data;
3. Monte Carlo simulation for the sum distribution.

They will be explained in detail in the following subsections.

4.1 The lagged synchronization for path travel time aggregation

For different links and trips, the dependence structure definition accounts for the time-lag along the trip, as different geometric relationship brings different time lag. We therefore consider the dependence is defined by considering geometrical factors:

1. The competing links, links which start at the same node and the entering time to which are the same: the dependence structure should be defined and estimated by a lag 0.
Definition 6. The correlation between two random travel time.

\[ \text{Cov}^{XY} = \frac{E(X(t)Y(t)) - E(X(t))E(Y(t))}{\sigma_X \sigma_Y} \]

Definition 7. The copula between the random travel time of two links.

\[ C^{XY}_t(x, y) = \left\{ (a, b) : F^{XY}_t(x, y) = C(F^X_t(a), F^Y_t(b)) \right\} \]

\[ C^{T_1...T_n}(t_1, \ldots, t_n) = \left\{ (x_1, \ldots, x_n) : F^{T_1...T_n}_t(t_1, \ldots, t_n) = C(F^X_1(t_1), \ldots, F^X_{t+s-1}(t_n)) \right\} \]

2. The succeeding links: the links along a trip, the entering time to which are increasing:

we define the lagged copula and lagged correlation as follows:

Definition 8. The lagged correlation between two random travel time.

\[ \text{Cov}^{XY}(s) = \frac{E(X(t)Y(t+s)) - E(X(t))E(Y(t+s))}{\sigma_X \sigma_{t+s}} \]

Definition 9. The lagged copula between the random travel time of two links.

\[ C^{XY}_{t,t+s}(x, y) = \left\{ (a, b) : F^{XY}_{t,t+s}(x, y) = C(F^X_t(a), F^Y_{t+s}(b)) \right\} \]

\[ C^{T_1...T_n,t,t+s}(t_1, \ldots, t_n) = \left\{ (x_1, \ldots, x_n) : F^{T_1...T_n}_{t,t+s}(t_1, \ldots, t_n) = C(F^X_1(t_1), \ldots, F^X_{t+s-1}(t_n)) \right\} \]

\[ C^{T_1...T_n}(t_1, \ldots, t_n) = \left\{ (x_1, \ldots, x_n) : F^{T_1...T_n}_{t,s+1,...,t+s-1}(t_1, \ldots, t_n) = C(F^X_1(t_1), \ldots, F^X_{t+s-1}(t_n)) \right\} \]

Where \( s_n \) is the series of specified time instant for links in a consecutive path. We set \( s(n) = \sum_{i=1}^{n-1} E(T_{s_i}). \)

The definitions here is to reflex the time varying dependence of the two related travel time processes. The time difference is defined as the expected difference of starting time instant on each arc under consideration while traveling through them in a path. And the two arcs starts from the same nodes, the lagged structure reduce to normal dependence structure as the time lag is 0.

3. For competing links which do start not start from the same node. That is, for two arcs \( AB \) and \( CD \), we select a reference node \( O \), and calculate the mean travel time for the path \( OA \) and \( OC \), Then the lag is estimated as \( ET_{OA} - ET_{OC} \)
The definition above specifies the way to organize data. For paths, we select the second synchronization scheme as it is specified in previous sections. Then when doing estimation based on this lagged dependent structure, the observations obtained at the corresponding time intervals will form the observation vector $X$. After the vector is formed, similar estimation procedure will be carried out for the whole vector $X$.

4.2 Lasso estimation of the lagged gaussian copula

After we specify the type of the dependent structure model, we go for the parameter estimation. The goal of the procedure designed in this section is to give a method which can work for any path with different length and limited data.

The challenge here is that for a long path, the data in common with all the links might be limited. The methodology we proposed is to aggregate the link travel time based on lagged gaussian copula, with covariance matrix adjusted by the lasso method. We consider the whole covariance matrix $P$ together, and estimate the inverse of covariance matrix directly by maximizing an objective function penalizing by the L-1 norm of $P^{-1}$ by Lasso [8].

1. The gaussian copula parameter estimation based on the monotonic transformation is given by:

$$\Sigma = \frac{1}{n}Z^T Z - \frac{1}{n}Z^T 11^T Z$$

where

$Z = \Psi^{-1}(F(X))$

$F = (F_1, F_2, \ldots, F_d)$ is the known marginal distribution

and

According to the definition of joint Gaussian distribution, it is Gaussian marginal distributions composed by a Gaussian copula. As we are considering the non-normal marginal distributions we take the monotone transform on data to do parameter estimation of gaussian copula. After the transformation, the data subject to a joint normal distribution with standard Gaussian margins.

2. Then the estimation of the covariance matrix for the Gaussian copula is conducted, the challenges here are,

(a) $n \sim p$, $\Sigma$ cannot be considered a good estimate

(b) $n \preceq p$, the empirical covariance $S$ is singular.
Using graphic Lasso method, we can generate a sparse, invertible estimate covariance matrix by solving the following problem:[2]and [1]:

$$
\max \log(\det X) - tr(\Sigma^T X) - v||X||_1
$$

where $|X|_1$ is the sum of the terms in $X$.

And the dual problem of this problem can is as follows:

As shown in , $L^1$ norm and $L^\infty$ are dual norm to each other

$|X|_1 = sup(X^TU||U|_\infty \leq 1$)

And then

$$
\max_{X \succeq 0} \log(\det X) - tr(\Sigma^T X) - v||X||_1 = \max_{X \succeq 0} \log(\det Y) - tr(\Sigma^T (X)) - v\min_{|U|_\infty \leq 1} (tr(X^TU))
$$

Exchange min and max to get the dual problem

$$
\min_{|U|_\infty \leq 1} \max_{Y \succeq 0} \{\log(\det Y) - tr(\Sigma^T (Y)) - vtr(Y^TU)\}
$$

and for the inner problem the derivative is

$$
Y^{-T} - \Sigma - vU = 0
$$

and

$$
Y = (\Sigma + vU)^{-1}
$$

here we use the fact the matrixes are symmetric

And inner problem has the solution

$$
-\log \det(\Sigma + vU) - p
$$

so the problem is then

$$
\min_{|U|_\infty \leq 1} -\log\det(\Sigma + vU) - p
$$
We just need to maximize

$$\min_{|W - \Sigma|_\infty \leq v} \log det(W)$$

where \( W = \Sigma + vU \) In language form it is

$$\max \log det(W) + \lambda(|W - \Sigma|_\infty - v)$$

It can be solved in a Block Coordinate Descent Algorithm as indicated in [1] In a series of problem in the form of

$$\min y^T W^{-1}_{j,j} y + \lambda(|y - \Sigma_j|_\infty - v)$$

where \( W_{j,k} \) denotes the matrix produced by removing row \( k \) and column \( j \). Let \( \Sigma_j \) denotes the column \( j \) with the diagonal element \( \Sigma_{jj} \) removed.

and for each such sub problem the dual problem is:

$$\min y^T W_{j,j} y - S_j y + v||y||_1$$

It is a L-1 constrained problem called lasso. We can then use the solver of lasso problem to get the solution.

This procedure achieves the following goals

1. The \( X \) is the inverse covariance matrix of the transferred data. And it is the parameter matrix for the gaussian copula.

2. The \( X^{-1} \) is always positive definite while the initial sample covariance matrix \( \Sigma \) is not necessarily when the number of common observations are smaller than the dimension.

3. The change of the parameter \( v \) change the scarcity of the estimated covariance matrix. Then we can make trade off on the dependence between links far away from each other if data is limited for estimation. That is, there is little data for the pair and the links are far from each other, the covariance can be adjusted to near 0 while keeping the covariance matrix positive definite.
4.3 Aggregation of the link travel time distribution and Simulation

The lasso modification enables the design to be invariant to the number of observations. And it is a promising way in aggregating link travel time based on sparse floating car data. After the estimation of the dependent structure, the choice of marginal distribution to generate the joint distribution is of several possible designs:

1. Based the copula method we can simulate the normal random variables according to the following algorithm: simulated the $N$ dimensional data from Gaussian joint distribution with covariance matrix $\Sigma$

2. Monotone changes from normal variables into a specific marginal distribution $F_i^{-1}(\Psi_i(X_i))$. There are several ways to select $F^{-1}$, if we consider the timely changes of marginal distribution, we consider

   (a) 1 All $T_i$ are identical in marginal distribution, although there dependent structure is governed by the estimated gaussian copula above.

   (b) 2 Each $T_i$ subject to a different distribution for different time instant.

   (c) 3 We consider to approximate the projected two dimensional distribution of $X_1$ and $X_j$ with the lagged bb1 copula, used in [23], and use the conditional distribution as the marginal distribution to do the aggregation. there are further two ways to do this

      i. 3.1 use the identical marginal distribution as in 1;

      ii. 3.2 use the timely marginal distribution as in 2;

3. Sum the marginal to get the path travel time according to the estimated copula and the marginal distribution

We interpret the three methods as follows: Method 1 is to reconstruct the joint distribution based on the lagged copula and the overall marginal distribution; Method 2 is to reconstruct the joint distribution based on the lagged copula and the timely marginal distribution on each link, it should give a timely approximation as it considers the time change of travel time. The drawback is the expectation of timely marginal distributions may not strictly satisfy (8) due to data limitation. Method 3 direct estimates the unknown marginal distribution by reconstructing the low dimensional joint distribution (which is the projection of the total joint distribution on the two dimensional space), and then compose them with the estimated high dimensional copula. The conditional distribution we considered in the paper is to condition on the entering time on a sequence of time. But we can go on to conditioning on both this time lag
and a latest observation on the first link for travelers who are trying to make choices at the exit of the first link. Method 3 provide such improvement.

In all, Method 1 is strict in both copula theory and lag vector definition. Method 2 is also strict reconstruction of joint distribution according to the copula theory while there is approximation as the timely mean travel time on each link may not satisfy (8) strictly. An further improvement to Method 2 is to discriminate the timely link travel time distribution into sub distributions conditioning on the short term mean $F_{t,ET}^X$. Then we choose the specific sub distribution to do the aggregation. This can be as strict as Method 1 while adding more flexibility in time. Method 3 is an approximation of joint distribution given the new observation on link 1 and it subject to both errors on not strict joint distribution and not strict time lag match.

Next, Monte Carlo simulation for the sum distribution is conducted based on the estimation to get the sum distribution of path travel time. And simulation of the path travel time can be conducted by the following procedure:

1. Diagonalize the estimated covariance matrix. as $\Sigma = B^t \Lambda B$
2. for $j = 1$ to $m$ and repeat the following procedures for $m$ times:
3. Generate the n independent normal variables $X$ with $\Psi(0, 1)$ and define $i$ as the random index in the set $1...n$;
4. Generate correlated normal variables as $Y = BX$
5. Take monotone transfer, as $Z_i = F_i^{-1}(\Psi_i(Y_i))$, as $F_i$ are selected by the marginal distribution discussed above, if necessary, we delete those data which is less than the shortest possible travel time.
6. define $Y(\omega_j) = \sum Z_i(\omega_j)$ as a realization of the stochastic path travel time.

This procedure is an alternative to the multidimensional integration for this complex high dimensional joint distribution. The experiments are carried out as validation in the experiment session.

### 4.4 Error estimation

Recall the main procedure for one scenario is to approximate $\sum_{i=1}^N T_{s_i}$ with $\sum_{i=1}^N T_{t_i}$ (8) and estimate $\sum_{i=1}^N T_{t_i}$ based on copula method. We consider the sources of error in the estimation for a scenario in this session.
1. We consider the error below.

\[
\sum_{i=1}^{N} T_{s_i} - \sum_{i=1}^{N} T_{t_i} = \sum_{i=1}^{N} (T_{s_i} - T_{t_i})
\]

**Theorem 3.** Define:

\[K_i(x)\]

such that

\[
\int P(T_{s_i} = x|s_i = a)P(s_i = a)da = K_i(x)P(T_{s_i} = x|s_i = t_i)P(s_i = t_i)(t_i - mins_i)
\]

\[k_i = \min_x K_i(x)\]

and

\[K_i = \max_x K_i(x)\]

\[\Delta_L^i = E(T_{t_i})(k_i P(s_i = t_i)(t_i - mins_i) - 1)\]

\[\Delta_U^i = E(T_{t_i})(K_i P(s_i = t_i)(t_i - mins_i) - 1)\]

\[\Gamma_L^i = E(T_{t_i}^2)(k_i P(s_i = t_i)(t_i - mins_i) - 1)\]

\[\Gamma_U^i = E(T_{t_i}^2)(K_i P(s_i = t_i)(t_i - mins_i) - 1)\]

then the total mean error of the estimation is bounded by

\[
\sum_{i=1}^{N} \Delta_L^i \leq \sum_{i=1}^{n} (ET_{s_i} - ET_{t_i}) \leq \sum_{i=1}^{N} \Delta_U^i
\]

the error of each marginal variance is bounded by

\[\Gamma_L^i - \Delta_U^i (ET_{s_i} + ET_{t_i}) \leq \Var T_{s_i} - \Var T_{t_i} \leq \Gamma_U^i - \Delta_L^i (ET_{s_i} + ET_{t_i})\]

for each \(i\)
An iterative procedure can be conducted for the estimation of error above, in the process we start from the 2-dimensional case to estimate \( \Delta_i \). For each \( \Delta_i \), we need the distribution of \( s_i \), \( T_{ti} \) and the parameters \( k_i, K_i \).

(a) The distribution of \( s_i \) is the path travel time distribution for the first 1 to \( i - 1 \) links.

(b) The distribution of \( T_{ti} \) can be estimated through kernel methods it is the marginal distribution used in the procedure in this paper.

(c) A harder problem is to get a suitable estimates for \( k_i \) and \( K_i \). Simulation based analysis can be conducted in future research.

2. 2 The normal copula assumption might not be a good fit for the actual underlying copula, Other copula family or nonparametric method can be used in future research.

3. 3 The parameter estimation of the copula covariance matrix is determined by the available data although the whole procedure works for limited data. According to Fan (2008)[6], the error of the estimated covariance matrix is of the order \( O(1/\sqrt{n}) \) (in Matrix infinity norm).

We then will make decisions by the decision rules introduced by Wan (2009)[23], including: mean variance decision rule, Transferred Mean Variance decision rule, First-order stochastic domination (FSD), second-order stochastic domination (SSD), area ratio rule, expected utility and Value-at-risk rule. And the path with both smaller mean \( \mu \) and variance \( \sigma^2 \), with smaller value of \( \mu + r\sigma \), which is dominated in the sense of FSD and SSD, with a larger area ratio, with a larger expected utility value, or with a smaller Value-of-risk, will be preferred.

## 5 Numerical examples

In this section, we first illustrate the basic procedure by estimate the overall path travel time distribution by one conditional path travel time distribution which satisfies (8). Routing decision are then made based on estimated distribution. A further aggregation of different scenario is given.

For the experiment, we select one twelve-link path between Allen town and Clinton through Highway 78 to do the estimation of path travel time while select two competing pathes near philadelphia to illustration of decision making process, the network is shown in Figure 1.

[ Insert Figure 1 here]
First, we change $\Delta$ the tolerance in time interval to find out a good discrete scale, fixing $v = 0.001$ and marginal distribution as the overall marginal link travel time distributions. The changes of fitting is shown in Figure 2, the graph shows the interval width is about $2\sigma$ the error is smaller. Too small a interval width makes the estimation go to specific path travel time experience and also subject to theoretical error, we will set $\Delta = 3\sigma$ in later experiment.

Second, we change different marginal distributions to do estimation while fixing $\Delta = 2\sigma$ and $v = 0.001$. According to the Figure 3 and distance statistics later, the overall marginal distributions yields less error in estimation, timely marginal distribution is flexible but subject to the limitation of data, hence brings more error. We will set marginal distributions as the overall marginal distributions in later experiments.

Third, we change Lasso penalty to do estimation while fixing $\Delta = 3\sigma$ and marginal distributions as the overall marginal link travel time distributions. The estimates based on different lasso penalties shrinkage between the totally independent assumption and the empirical estimation, as shown in Figure 4. The more penalty used, the more the estimated covariance structure tend to be independent and the worse the approximation is. This shows that the dependence between links is heavy and independent assumption does not work. Generally speaking, little penalty yields better estimation, and we will set $v$ as the smallest value 0.001 for accuracy, in later experiments.
[ Insert Figure 4 here]

Fourth, we then conduct a weighted sum of limited different scenarios, the scenarios are picked by setting $\Delta = 3\sigma$ fixed and changing the lag vector $V$. We take two other scenario where the travel time of the tenth link is within $\mu_10 + 1.5 \cdot \sigma_10$, $\mu_10 + 4.5 \cdot \sigma_10$ and the later link travel time are as before. Then the lag vector is $V_2 = [t_1, \ldots, t_10, t_11 + 3\sigma_10, t_12 + 3\sigma_10]$ and here, $t_i$ as defined in (8). $\sigma_10 = \max_i \sigma_i$, so the two scenario will not intersect with each other. The relative probability of the two scenario is approximately estimated as 4.5:1, by applying Chebyshev’s inequality to the travel time distribution of Link 1, then the new estimation for the path travel time is $P(x|\Delta) = \frac{4 \cdot 5}{5.5} P(x|V_1, \Delta) + \frac{1}{5.5} P(x|V_2, \Delta)$, and the estimates yields larger P-value in 1. We can see the estimation is better than the one only base on the main scenario. Actually, the more discrete scenario one take, the better the final estimation is. The continuous approximation within the discrete scheme is an efficient and flexible way for path travel time distribution estimation.

We calculate the distance of estimated distribution from the empirical path travel time distribution, as shown in Table 1

Finally, to show the comparison of the two competing pathes, we use the overall marginal distribution with lasso penalty $v = 0.001, \Delta = 3\sigma$, and compare the two
paths by their major scenarios. Based on our methods, we estimate the path travel time distribution as in figure and the decision statistics is calculated in Table 2. The path ends at different sides of a bridge, we do not count for the turning for the bridge as the data is no available for the section on the bridge. The estimated path travel time distribution is shown in the Fig ??.

[Insert Figure 5 here]

Then the decision statistics introduced in [23] are calculated to make routing decision based on the estimated path travel time distribution, as in 2.

Most decision rules will prefer Path 2. The transferred mean-variance prefers Path 1 as the selection of risk aversion $r$ gives more weight on variance (AB is preferred when $r$ takes value in $(0.0482, \infty)$). The two paths are not comparable under mean-variance decision rules and. Traveler can then select the statistics to make his own decision.

For long paths usually we do not have enough historical travel time observations, in this case there are just about 10 observations for either path while hundreds more observations for each link. The direct aggregation scheme based on link data saves the cost of state enumeration and enables the decision based on travel time distribution for realtime systems.
6 Conclusion and Future research

In the paper, we propose a continuous approximation for the conditional path travel time distribution within a discrete state space. For each scenario characterized by a lag vector $V$ and time tolerance $\Delta$, a Gaussian copula based estimator is proposed with Lasso method to estimate the copula parameter matrix. The whole estimation framework, the introduction of copula method to path travel time distribution and the introduction of Lasso for Gaussian copula estimation are the new development both in theory and applications.

With respect to the decision making aspect, non-normal dependent path travel time can be estimated based on floating car data and stochastic routing decision can be made according appropriate decision rules. Decision rules are shown and given interpretation in the transportation context, which make it possible to a better objective decision for the travelers.

In future research, the challenges mainly lie in:

1. $\Delta$ acts as the trade off the specific realization of path travel time ($\text{when} \Delta = 0$) and the conditional path travel time distribution in the scenario ($\text{when} \Delta > 0$).

The larger $\Delta$ is there is more theoretical error as we assume probability constancy in longer time while more data to do estimation so less data error. An optimal
Table 1: Distance measures for different estimation

<table>
<thead>
<tr>
<th>Δ</th>
<th>2σ</th>
<th>3σ</th>
<th>4σ</th>
<th>6σ</th>
<th>Combine Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2</td>
<td>0.0115</td>
<td>0.0155</td>
<td>0.0111</td>
<td>0.0133</td>
<td>0.0154</td>
</tr>
<tr>
<td>P-value</td>
<td>1.43e-008</td>
<td>1.78e-007</td>
<td>6.16e-008</td>
<td>3.43e-008</td>
<td>1.2955e-007</td>
</tr>
<tr>
<td>Marginal</td>
<td>Overall</td>
<td>Realtime</td>
<td>Conditional 1</td>
<td>Conditional 2</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>0.0176</td>
<td>0.0320</td>
<td>0.0782</td>
<td>0.0275</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>1.70e-008</td>
<td>3.51e-021</td>
<td>1.52e-055</td>
<td>1.36e-017</td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td>v = 0.001</td>
<td>v = 0.01</td>
<td>v = 0.05</td>
<td>v = 0.1</td>
<td>v = 0.5</td>
</tr>
<tr>
<td>L2</td>
<td>0.0176</td>
<td>0.0131</td>
<td>0.0171</td>
<td>0.0189</td>
<td>0.0125</td>
</tr>
<tr>
<td>P-value</td>
<td>1.70e-008</td>
<td>3.80e-008</td>
<td>7.61e-009</td>
<td>3.08e-008</td>
<td>7.5282e-009</td>
</tr>
</tbody>
</table>

Table 2: Decision statistics for different rules

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(μ, σ)</th>
<th>μ + rσ²</th>
<th>FSD first violation</th>
<th>SSD first violation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path 1</td>
<td>(1199,130.5)</td>
<td>2051</td>
<td>∞</td>
<td>∞</td>
</tr>
<tr>
<td>Path 2</td>
<td>(993.8,145.9)</td>
<td>2058</td>
<td>739.4</td>
<td>0</td>
</tr>
<tr>
<td>Preference</td>
<td>None</td>
<td>Path 1</td>
<td>Path 2</td>
<td>Path 2</td>
</tr>
</tbody>
</table>

Exponential utility
a=-1/1000  AR   tU=1600  VAR(Quantile)  5%

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Path 1</th>
<th>Path 2</th>
<th>Path 2</th>
<th>Path 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path 1</td>
<td>0.2476</td>
<td>251.5</td>
<td>1832</td>
<td></td>
</tr>
<tr>
<td>Path 2</td>
<td>0.3509</td>
<td>559.5</td>
<td>1394</td>
<td></td>
</tr>
</tbody>
</table>

value for Δ is in need.

2. More data is in need for the research. As the scheme includes a detailed division of link travel time distribution based on the entering time and the short term mean (related to different scenarios). More dense data set for vehicle travel time is in need for further validation.

3. Other copula family or nonparametric method for copula can be studied. The gaussian copula is restricted by its specific parametric form and may not be able to fully describe the high dimensional dependent structure. Kernel method can be used.

4. New Risk measures are introduced in this paper for the travelers in a stochastic context. Comparison of these different risk measures for the stochastic routing decision need to be conducted in future. Meanwhile, different decision rules will lead to different system behavior, the associated equilibrium models need to be studied.
7 Proof

Proof of Theorem 6:

\[ \int P(T_{s_2} = x | T_{s_1} = a)P(T_{s_1} = a)da = K_2(x)P(T_{s_2} = x | T_{s_1} = ET_{s_1})P(T_{s_1} = ET_{s_1})(ET_{t_1} - minT_{t_1}) = K_2(x)P(T_{s_2} = x | T_{s_1} = ET_{t_1})P(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{t_1}) \]

\(K_2(x)\) is a modifying constant determined by the distribution \(s_2 = s_1 + T_{s_1} = T_{s_1}\) such that

\[ \int P(T_{s_2} = x | T_{s_1} = a)P(T_{s_1} = a)da = K_2(x)P(T_{s_2} = x | T_{s_1} = ET_{s_1})P(T_{s_1} = ET_{t_1})(ET_{s_1} - minT_{s_1}) \]

since \(s_1 = t_1\), it is therefore the \(K_2(x)\) such that
\[ \int P(T_{s_2} = x|T_{s_1} = a)P(s_2 = a)da = K_2(x)P(T_{s_2} = x|s_2 = t_2)P(s_2 = t_2)(t_2 - mins_2) \]

Then

\[
ET_{s_2} = \int xK_2(x)P(T_{s_2} = x|T_{s_1} = ET_{t_1})P(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{t_1})dx \\
= \int xP(T_{s_2} = x|X_{s_1} = ET_{t_1})K_2(x)dxP(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{t_1})
\]

As when \( s_1 = t_1 \) and \( T_{s_1} = ET_{s_1} = ET_{t_1} \) we have

\[ s_2 = s_1 + T_{s_1} = s_1 + ET_{s_1} = t_1 + ET_{s_1} = t_2 \]

we have

\[ P(T_{s_2} = x|T_{s_1} = ET_{t_1}) = P(T_{t_2} = x) \]

and

\[ ET_{s_2} = \int xK_2(x)P(T_{t_2} = x)dxP(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{t_1}) \]

Take \( K_2 = \max_x(K_2(x)) \)

\[ \Delta_2 = ET_{s_2} - ET_{t_2} \]
\[ = \int xK_2(x)P(T_{t_2} = x)dxP(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{s_1}) - \int xP(T_{t_2} = x)dx \]
\[ \leq \int xP(T_{t_2} = x)dx(K_2P(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{s_1}) - 1) \]
\[ = E(T_{t_2})(K_2P(T_{s_1} = ET_{t_1})(ET_{t_1} - minT_{s_1}) - 1) \]
\[ = E(T_{t_2})(K_2P(s_2 = t_2)(t_2 - mins_2) - 1) \]

Similarly, for lower bound

\[ \Delta_1 \geq E(T_{t_2})(k_2P(s_2 = t_2)(t_2 - mins_2) - 1) \]

where \( k_i = \max_x K_i(x) \)

similarly when \( n=i, \)
\[
\int P(T_{s_i} = x | s_i = a) P(s_i = a) da = K_i(x) P(T_{s_i} = x | s_i = t_i) P(s_i = t_i) (t_i - mins_i)
\]

since
\[
P(T_{s_i} = x | s_i = t_1 + \sum_{j=1}^{i-1} ET_{t_j}) = P(T_{t_i} = x)
\]
as we have \(s_i = s_1 + \sum_{j=1}^{i-1} T_{s_j} = t_1 + \sum_{j=1}^{i-1} ET_{t_j} = t_i\)

\[
ET_{s_i} = \int K_i(x) P(T_{s_i} = x | s_i = t_i) P(s_i = t_i) (t_i - mins_i) x dx
\]
\[
= \int xK_i(x) P(T_{t_i} = x) dx P(s_i = t_i) (t_i - mins_i)
\]

And then take \(K_i = \max_x K_i(x)\)

\[
\Delta_i = ET_{s_i} - ET_{t_i}
\]
\[
= \int xK_i(x) P(T_{t_i} = x) dx P(s_i = t_i) (t_i - mins_i) - \int xP(T_{t_i} = x) dx
\]
\[
\leq E(T_{s_i}|s_i = t_i)(K_i P(s_i = t_i) (t_i - mins_i) - 1)
\]
\[
= E(T_{t_i})(K_i P(s_i = t_i) (t_i - mins_i) - 1)
\]
\[
\equiv \Delta_i^L
\]

Similarly, for lower bound

\[
\Delta_i = ET_{s_i} - ET_{t_i}
\]
\[
\geq E(T_{t_i})(k_i P(s_i = t_i) (t_i - mins_i) - 1)
\]
\[
\equiv \Delta_i^U
\]

where \(k_i = \max_x K_i(x)\)

The sum of \(\Delta_i\) can be used as modification to the initial estimate.
\[
\sum_{i=1}^{n} ET_{s_i} = \sum_{i=1}^{n} ET_{t_i} + \sum_{i=1}^{n} \Delta_i
\]
and
\[
\sum_{i}^{N} \Delta_i^L \leq \sum_{i=1}^{n} (ET_{s_i} - ET_{t_i}) \leq \sum_{i}^{N} \Delta_i^U
\]

Similarly for the variance, there is a bound of error in each of the \(T_{s_i}\). Consider

\[
\Gamma_i = ET_{s_i}^2 - ET_{t_i}^2
= \int x^2 K_i(x) P(T_{t_i} = x) dx P(s_i = t_i)(t_i - mins_i) - \int x^2 P(T_{t_i} = x) dx
\]

So
\[
\Gamma_i^L \leq \Gamma_i \leq \Gamma_i^U
\]
where
\[
\Gamma_i^L = E(T_{t_i}^2)(K_i P(s_i = t_i)(t_i - mins_i) - 1)
\]
\[
\Gamma_i^U = E(T_{t_i}^2)(K_i P(s_i = t_i)(t_i - mins_i) - 1)
\]

And we have
\[
VarT_{s_i} - VarT_{t_i} = ET_{s_i}^2 - (ET_{s_i})^2 - (ET_{t_i}^2 - (ET_{t_i})^2)
= \Gamma_i - \Delta_i(ET_{s_i} + ET_{t_i})
\]
So
\[
\Gamma_i^L - \Delta_i^U(ET_{s_i} + ET_{t_i}) \leq VarT_{s_i} - VarT_{t_i} \leq \Gamma_i^U - \Delta_i^L(ET_{s_i} + ET_{t_i})
\]
for each \(i\)

Q.E.D

References

Gaussian graphical models. In Proceedings of the 23rd international conference on

Vehicle Data: Link-based vs. Path-based. Transportation Research Record, pages

[4] CF Daganzo. The cell transmission model: a dynamic representation of high-
way traffic consistent with the hydrodynamic theory. Transportation research.

[5] DJ Dailey, ZR Wall, SD Maclean, and FW Cathey. An algorithm and implementa-
tion to predict the arrival of transit vehicles. In Intelligent Transportation


and Kendall’s tau for pairs of continuous random variables. Journal of Statistical


[9] L. Fu and LR Rilett. Expected shortest paths in dynamic and stochastic traffic


2006.

[14] X. Nie and HM Zhang. Delay-function-based link models: their properties and


