

断片的なプローブ軌跡の接合による区間旅行時間の期待値と分散の推定*

Estimation of Expectation and Variance for a Section Travel Time based on Fragmentary Probe Trajectories*

上杉友一**, 井料隆雅***, 小根山裕之****, 堀口良太*****, 桑原雅夫*****

By Yuichi UESUGI**, Takamasa IRYO***, Hiroyuki ONEYAMA****, Ryota HORIGUCHI*****, Masao KUWAHARA*****

1. はじめに

本稿は、プローブ車両から得られるリンク旅行時間情報を用いて、ある対象区間の旅行時間を推定する際の、対象とするプローブ軌跡の選択方法と期待値・分散の推定精度の関係を検証するものである。

現在のプローブデータを用いた旅行時間推定の可否は、プローブ車両の台数に大きく依存していると言える。特に、旅行時間の推定対象となる区間が長い場合には、対象時間内に区間全体を完全に通過するプローブ車両の台数は、必然的に減少してしまう。

一方、情報利用者のニーズで決定される情報提供区間と、自由意志に基づいて行動するプローブ車両の通過区間は、必ずしも一致しない。また、タクシーのようにある属性をもった車両をプローブ車両として採用する場合、プローブ軌跡にもある程度の偏りが生じると考えられる。そのため、ニーズに基づく対象区間を完全に通過するプローブ軌跡は、ごく少数しか存在しない、あるいは全く存在しない、という状況が容易に想像できる。

実際に、平成12年と平成13年に横浜市域で実施されたプローブ実験で得られたデータ⁽¹⁾をみても、ある属性を持った車両をプローブとして採用した場合には、プローブの経路の偏りとサンプルの不足が生じる、という問題を確認することが出来る。

そのため、対象区間を完全に通過する車両だけでなく、対象区間を部分的に通過する車両(図1参照)の旅行時間も推定に利用することで、推定に

利用可能な旅行時間データの数を増やすことが望まれる。

また、ユーザーへ提供すべきサービスという側面から考えると、対象区間全体の旅行時間の期待値だけではなく、旅行時間の変動幅や信頼区間のような情報も提供すべきであると考えられる。

そこで本稿では、シミュレーションにより再現された交通状況において、対象区間を完全に通過するプローブ軌跡を用いて、対象区間全体の旅行時間の期待値と分散を推定した場合と、対象区間を部分的に通過するプローブ軌跡も加えて、対象区間全体の旅行時間の期待値と分散を推定した場合で、どの程度の違いがあるかを検証した。さらに、断片的なプローブ軌跡を利用する際に、後の2章で述べる交通工学的な解釈・処理を加えることにより、区間全体の旅行時間期待値と分散の推定精度がどの程度向上するか、という点についても検証を行った。

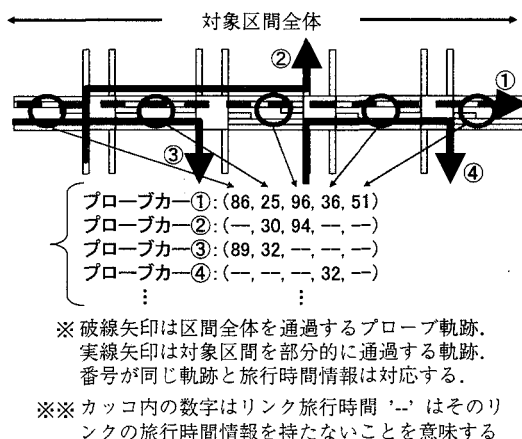


図1: 完全又は断片的な旅行時間情報を持つプローブ軌跡

2. 複数リンク間の旅行時間と分散の推定方法

本章では、対象区間内の複数リンクにおける各リンクの旅行時間と分散から、区間全体の旅行時間の期待値と分散を算出する方法について述べる。

*キーワード: 交通情報, ITS, プローブ

**学生員, 東京大学大学院工学系研究科

連絡先: 〒153-8505 東京都目黒区駒場 4-6-1

東京大学生産技術研究所 桑原研究室

***正会員, 工博, 東京大学国際産学共同研究センター

****正会員, 東京大学生産技術研究所

*****正会員, 工博, 株式会社アイ・トランスポート・ラボ

*****正会員, Ph.D., 東京大学生産技術研究所

ある区間において、ユーザーへの旅行時間情報提供を考える時、その推定がプローブデータを用いて行うものである場合には、前章で述べたサンプルの不足や属性による偏りなどという理由から、対象区間の通過車両が存在しないため旅行時間の推定が不可能である、もしくは可能であってもサンプル不足のため精度が低い、という問題が考えられる。そのため本章では、対象区間の推定旅行時間を、区間内の各リンクの旅行時間情報を積み上げて推定する、という手法を考えることにする。

(1) リンク旅行時間からの区間旅行時間の推定
プローブ車両は、各リンクに進入した時と退出した直後にそれぞれ時刻を取得しており、その情報からリンク旅行時間を算出することができる。その対象区間内の各リンク旅行時間から、対象区間全体の、旅行時間の期待値と分散を求める際には、次に示す式(1)、(2)を用いる事ができる。

$$\mu = \sum_{i=1}^l \mu_i \quad \dots\dots(1)$$

$$s^2 = \sum_{i=1}^l s_i^2 + 2 \sum_{i>j} s_{ij} \quad \dots\dots(2)$$

μ : 対象区間全体の旅行時間の期待値
 μ_i : リンク*i*における旅行時間期待値
 s^2 : 対象区間全体の旅行時間の分散
 s_i^2 : リンク*i*における旅行時間分散
 s_{ij} : 対象区間内のリンク*i*と*j*の旅行時間共分散

例えば、ある1時間に対象区間を通過したプローブデータを用いれば、この期待値と分散は、その1時間内における対象区間全体の旅行時間の期待値と分散と言うことができる。

また、図1に示すように、多くのプローブ車両は、対象区間内のすべてのリンクに対して、完全な旅行時間情報を所持しているわけではない。

そこで、断片的な旅行時間情報しか持たない車両について、あたかもそれらの車両が対象区間を完全に通過したようにみなし、偶発的に対象区間中の部分的な旅行時間が欠落したと考えることで、全車両が対象区間を完全に通過した場合と同様に推定を行うことができる。すなわち、1 or 0 の確率変数を用いることで、あるプローブが所持しているはずである、対象区間内における全てのリンク旅行時間情報の中から、確率的にいくつかのリンク旅行時間にマスクをかけ、あたかもあるリンク

の旅行時間情報が欠落しているかのように表現するものとする。

以下には、断片的なプローブ旅行時間データを用いた場合における、各リンクの旅行時間の期待値・分散・共分散の計算方法を記す。

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{k=1}^m \delta_{ik} T_{ik} \quad \dots\dots(3)$$

$\hat{\mu}_i$: リンク*i*における旅行時間期待値の推定値

n_i : リンク*i*を通過するプローブ車両台数

T_{ik} : リンク*i*における車両*k*の旅行時間, $k=1,2,\dots,m$

$$\hat{s}_i^2 = \alpha \cdot \left\{ \frac{1}{n_i} \sum_{k=1}^m \delta_{ik} T_{ik}^2 - \left(\frac{1}{n_i} \sum_{k=1}^m \delta_{ik} T_{ik} \right)^2 \right\} \quad \dots\dots(4)$$

$$\alpha = \frac{N-1}{N} \cdot \frac{n_i}{n_i-1} \quad \text{: 補正項}$$

N : サンプルを行う母集団の車両数

$\hat{s}_i^2$: リンク*i*における旅行時間分散の推定値

$$\delta_{ik}: 0 \text{ or } 1 \text{ の確率変数 } \left(n_i = \sum_{k=1}^m \delta_{ik}, \quad n_{ij} = \sum_{k=1}^m \delta_{ik} \delta_{jk} \right)$$

$$\hat{s}_{ij} = \beta \cdot \left\{ \frac{1}{n_{ij}} \sum_{k=1}^m \delta_{ik} T_{ik} \delta_{jk} T_{jk} - \frac{1}{n_{ij}^2} \left(\sum_{k=1}^m \delta_{ik} T_{ik} \sum_{k=1}^m \delta_{jk} T_{jk} \right) \right\} \quad \dots\dots(5)$$

$$\beta = \frac{N-1}{N} \cdot \frac{n_{ij}}{n_{ij}-1} \quad \text{: 補正項}$$

$\hat{s}_{ij}$: リンク*i*とリンク*j*の旅行時間共分散の推定値

n_{ij} : リンク*i*と*j*の旅行時間を共に持つ車両台数

ここで、 α, β は、統計量を不偏推定量とするための補正項であり、その必要性及び導出の詳細については巻末の付録に付す。

また、前述の数式の補足をする、式(3)における δ_{ik} の意味は、車両 k がリンク i の旅行時間情報を所持しているかないかを、1 or 0 で確定する確率変数で表したものである。つまり、 δ_{ik} の k に関する総和は、リンク i を通過するプローブ車両台数 n_i に等しくなる。すなわち式(3)は、サンプルされたプローブ車両のうち、リンク i の旅行時間情報を所持するものだけを用いて、リンク i の旅行時間の期待値推定を行うものである。

式(5)についても同様の考え方で、サンプルされたプローブ車両のうち、リンク i とリンク j の両方の旅行時間情報を所持するものだけを用いて、リンク i とリンク j の旅行時間共分散の推定を行うものである。

(2) ケース設定

以下では区間全体の旅行時間の期待値 μ と分散 s^2 を推定する3つのアプローチを考える。

(a) ケース1

ケース1では、サンプリングされた車両の中から、対象区間を完全に通過する車両だけを抜き出して、推定を行う。

(b) ケース2

ケース2では、サンプリングされた車両の中から、対象区間を完全に通過する車両に加え、対象区間を部分的に通過する車両も考慮して推定を行う。

(c) ケース3

ケース3では、ケース2に交通工学的な解釈・処理を加えたものを考える。具体的には、ケース2の手法に基づいて、断片的な旅行時間情報を利用する際に、系統的に異なる旅行時間データを排除する作業を行う。

ケース3を考慮する交通工学的な視点について多少述べると、図2のように右左折で対象区間に流入し、右左折で対象区間から流出する車両の旅行時間は、対象区間内において、そのターンの直前・直後の旅行時間には、系統的な差異が生じると考えられる。その原因としては、右左折流入によるオフセットのずれやターンの際の遅れ時間などが考えられる。そのため、このような系統的な差異を含む旅行時間は、推定には用いるべきではないと考える。

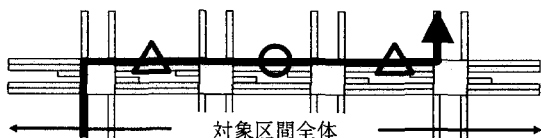


図2：ケース2とケース3の違い
(ケース3では○の旅行時間だけを使用)
(ケース2では△と○の両方を使用)

以上、3つのアプローチの概略を図3に示す。

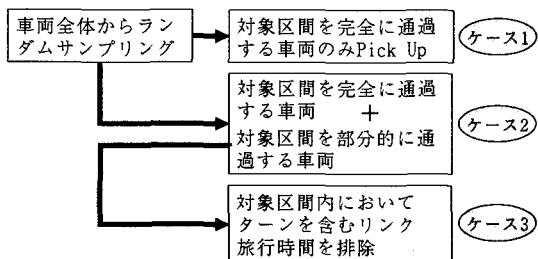


図3：3つのアプローチの概略

3. 計算機実験による検証

本章では、前章で述べた推定方法を用い、シミュレーションで再現された交通状況に対して旅行時間推定を行う手順について説明を付す。

まず、街路網の交通シミュレーション AVENUE を用いて、10 リンク 9 交差点から成る直線幹線道路（平均スパン 300m 程度）により構成されるネットワーク上で非飽和交通流（ボトルネック交差点飽和度 0.7 程度）を 1 時間再現した。

ネットワーク中の各 OD では、プローブ車両はランダムに分布していると考え、全車両データからプローブ車両の混入率に対応するサンプル率でランダムサンプリングを行い、プローブ車両を抽出する。この抽出されたプローブ車両の旅行時間情報を、前章で示した3つの手法を用いて、1 回の試行における対象区間全体の旅行時間の期待値と分散を推定する。

そして、前記の計算を十分な回数（1000 回）だけ繰り返し、対象区間全体の旅行時間の期待値と分散の分布形状を求め、その分布形をもとに最終的な旅行時間期待値と分散を算出する。

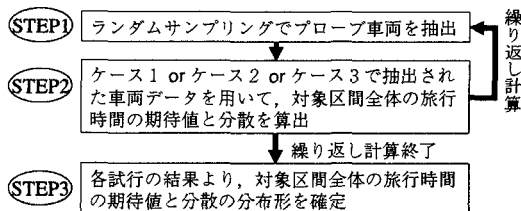


図4：計算機実験による検証手順

図4には、検証方法の手順の概略を示す。

また、ここで推定すべき値は、対象時間内の任意の時間に出発し、かつ、対象時間内に区間全体を通過した車両の旅行時間の期待値 μ と分散 s^2 であるが、これらを推定するのは困難であるため、ケース1におけるサンプル率 100% 時の、区間全体の旅行時間の期待値と分散をリファレンスとして、本検証を行うものとする。

4. 結果及び考察

本章では前章の手順と手法に基づいて、プローブ車両の混入率を変化させて、対象区間全体の旅行時間および旅行時間分散についての計算機実験を行った結果を示す。さらにその結果に対し、的中

率という指標を定義し、この指標を用いて各推定手法を評価し、考察を行うものとする。

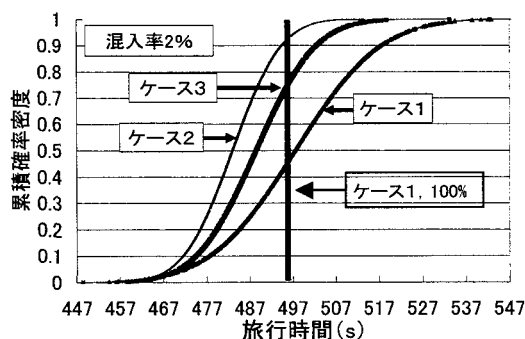


図5：混入率2%時の区間全体の旅行時間期待値分布

まず、プローブ混入率が2%および10%の時に、繰り返し試行1000回について、試行毎の対象区間全体の旅行時間期待値をソートして、累積確率密度の分布として表した図5、図6を示す。

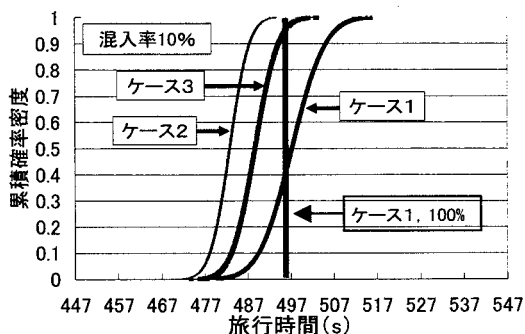


図6：混入率10%時の区間全体の旅行時間期待値分布

また、混入率が2%および10%の時、試行毎の、対象区間全体の旅行時間分散値の累積確率密度分布を示したものが図7、図8である。

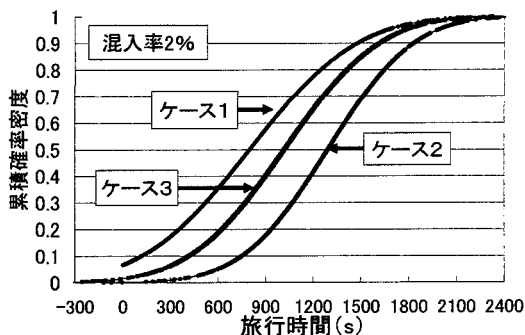


図7：混入率2%時の区間全体の旅行時間分散分布

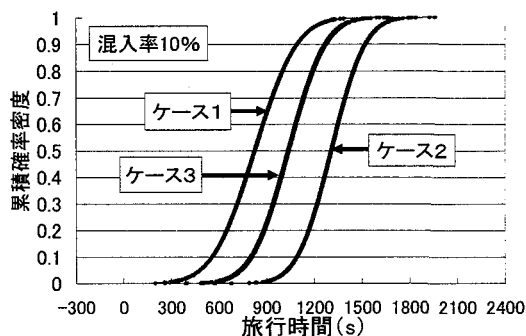


図8：混入率10%時の区間全体の旅行時間分散分布

図5、図6の区間旅行時間の期待値に注目すると、ケース1が最も大きく、次いでケース3、ケース2の順である。今回再現された交通状況では、右左折を含む旅行時間を排除した方が旅行時間期待値は長くなっているが、これはシミュレーション中のオフセット設定の関係で、直進でリンクを通過するよりも、系統の途中で従道路から右左折して主道路へ進入してくる車両の方が、当該リンクを短い時間で通過してしまうためである。

次に、図7、図8に示される区間旅行時間の分散値に注目すると、ケース2が最も大きく、次いでケース3、ケース1の順である。この結果から、ターンを含む旅行時間を排除した方が、区間旅行時間の分散を小さく出来ることがわかる。また図7に注目すると、ケース2およびケース3において対象区間全体の旅行時間分散を推定した場合、混入率が低い時のごく僅かな試行において、負の区間旅行時間分散が推定されている。この理由は、ケース2およびケース3において、本来、不完全な旅行時間情報しか所持していないプローブ軌跡に関しても、あたかも対象区間全体の完全な旅行時間情報を所持しているかのように推定を行った為である。そのため、旅行時間情報が完全な場合に比べて、各共分散の推定値を負側に大きく見積もってしまうと、式(2)により、分散共分散行列の成分の線形和で表現される区間分散推定値は結果的に負になる可能性が生じると考えられる。旅行時間情報が完全となるケース1の場合には、分散推定値が負になることは無い。

また、区間旅行時間の期待値と分散の両者について、プローブ混入率が2%の時と10%の時を比較すると、2%の累積分布形は10%のそれに比べ、なだらかである。このことは、混入率が低いと、提供できる区間旅行時間の変動幅が大きくなることを意味している。

次に、各ケースの評価として、的中率という指標を考える。的中率とは、例えば区間旅行時間期待値の推定値について、ケース1におけるプローブ混入率 100%時の区間旅行時間期待値を基準値とし、各ケースにおける1回の試行の推定値が、全試行回数中において、どの程度の割合で、基準値から $\pm 10(s)$ の範囲に収まるか、ということを示す指標であると定義する。また、区間旅行時間分散についても同様に、ケース1におけるプローブ混入率 100%時の区間旅行時間分散を基準値とし、各ケースにおける試行の推定値が、全試行回数中で基準値から $\pm 300(s^2)$ の範囲に収まる割合、と定義する。

図9には、 $\pm 10(s)$ 範囲の区間旅行時間の的中率を示す。

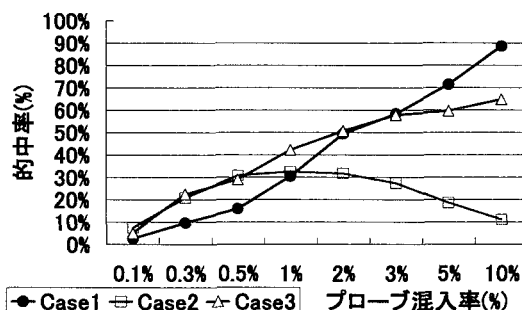


図9： $\pm 10(s)$ 範囲の区間旅行時間の期待値的中率

図9において、プローブ混入率が低い場合（1%以下の場合）には、ケース2とケース3の的中率が上回っている。しかしながら、混入率が3%を超えると、ケース1の的中率が上回ることが分かる。

この理由としては、ケース1では、混入率が低い時にはプローブ車両の絶対サンプル数が少ないため、精度の高い推定を行うことが難しくなる、という点が挙げられる。しかしながら、プローブ混入率がある程度上昇すると、プローブ車両数を確保することができるようになり、ケース1での推定精度が向上する一方で、部分的な区間しか通過しないプローブ車両の旅行時間情報を用いて区間旅行時間期待値と分散を推定するケース2とケース3では、系統誤差が推定値に与える影響を相対的に無視できなくなり、推定精度が低下する。

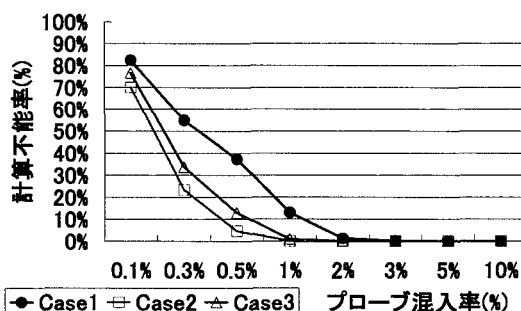


図10：区間旅行時間期待値が計算不能となる確率

図10には区間旅行時間の期待値が計算不能になる確率を示す。ここで計算不能になる確率について述べる。プローブ混入率が低い場合、各ケースにおいて、絶対プローブ車両数の少なさのため、対象区間の特定リンクを通過する車両が1台も存在しない、という状況が存在する。このようなリンクが対象区間に1リンクでも存在する場合、区間旅行時間期待値の推定は不可能であると判断し、計算不能として処理する。

また図10における計算不能率に注目すると、プローブ混入率が低い場合、計算不能となる確率はケース1が最も高い。

この理由は、ケース1では、対象区間を完全に通過するプローブ車両の旅行時間情報のみを用いて推定を行っているが、プローブ混入率が低い場合には、プローブ車両の絶対数の不足により対象区間を完全に通過する車両が皆無もしくは1台しか存在しない、という状況が容易に生じる。そのため、前述の計算不能率の定義に従うと、プローブ混入率が低い時には計算不能率が高くなる。しかしながら、今回の計算機実験においては、プローブ混入率が3%を超えると、各ケースにおいて計算不能となることは無くなる。このことは、ある程度以上のプローブ混入率においては、ケース1の推定精度が高くなる、という前述の結果に対応している。

すなわち、プローブ混入率が低い時は、計算不能を回避するため、ケース2およびケース3の手法が有効であるが、混入率がある程度高くなると、その的中率から判断して、ケース1の手法がもっとも適当であると考えられる。このことは、実際にプローブ実験を行って得られたデータを用いて旅行時間推定をする際に、そのプローブ車両の混入率によって、的中率の観点での最適な推定手法が異なることを意味している。

同様に、図 1 1 には $\pm 300(s^2)$ 範囲の区間旅行時間分散的中率を、図 1 2 には区間旅行時間分散の計算不能率を示す。

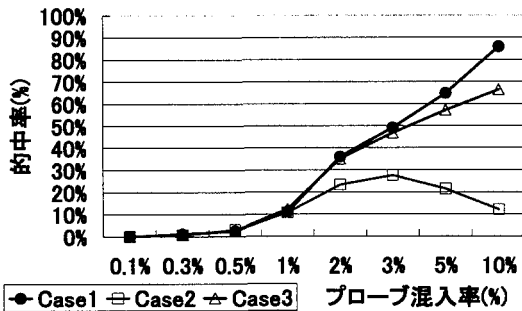


図 1 1 : $\pm 300(s^2)$ 範囲の区間旅行時間分散的中率

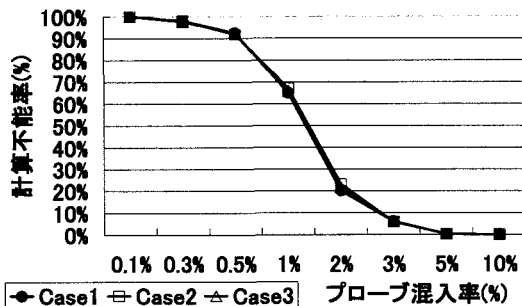


図 1 2 : 区間旅行時間分散が計算不能となる確率

区間旅行時間分散的中率に関して、プローブ混入率が低い時には、各ケースとも大きな差異は見られない。しかし、混入率が上昇すると、ケース 1 が最も高い中率を示し、それに続いてケース 3 の中率が高い値を示す。

この理由は、プローブ混入率が低い場合には計算不能率の高さからケース毎の差異はほとんど無いが、混入率が上昇すると、計算不能となる確率が低くなるため、系統的な誤差を含まないケース 1 の推定値が、結果的に最も中率が高くなると考えられる。

区間旅行時間分散の計算不能率について、各ケース間で混入率の違いによる差異は見られなかった。

また、ケース 2 における区間旅行時間期待値および分散的中率に関して、プローブ混入率が一定以上になると中率が低下している。この理由は、系統的に大きな差異を含む旅行時間も利用して区間全体の旅行時間期待値と分散を推定しているため、推定値に偏りが生じているためと考えられる。

以上をまとめると、プローブ混入率が低い時は、区間旅行時間期待値と分散の推定にはケース 2 と

ケース 3 の手法が有効である。混入率がある程度高くなると（今回の計算機実験では 3% を越えると）計算不能に陥る可能性がなくなるため、ケース 1 の手法を用いるのが有効であると言える。

5. 結論

本稿では、対象区間に対して断片的なプローブ軌跡から、対象区間全体の旅行時間の期待値と分散を推定する方法を検討した。その結果、プローブ車両の混入率が低い場合には、断片的なプローブ軌跡を用いる推定手法が有効であることが確認された。

また、本稿における推定手法は、プローブに限らず、AVI やビーコンといった車両を追跡できるセンサーで計測した旅行時間を繋げる際にも有効であるといえる。

【参考 URL】

街路網交通シミュレーションモデル AVENUE
<http://www.i-transportlab.jp/products/avenue>

【参考文献】

- (1) 自動車走行電子技術協会: ITS の社会的有効性向上に係るシステムの実証・報告書（平成 14 年 3 月）、2002

【付録】 2 章、式(4)、式(5)中の補正項について
 プローブ車両のサンプリング抽出は、本質的には非復元抽出であるため、プローブ車両から得られる区間旅行時間の分散の推定量を不偏推定量として計算する場合には、復元抽出を前提としている一般的な公式とは異なった式を考慮する必要がある。

いま N 台の車両から非復元抽出で n 台の車両をサンプリングし、分散を算出することを考える。その際、区間全体の旅行時間分散の真値 σ^2 と、全プローブ車両における分散値 s_p^2 は次のように書ける。

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N X_i^2 - \left(\frac{1}{N} \sum_{i=1}^N X_i \right)^2$$

$$s_p^2 = \frac{1}{n} \sum_{i=1}^n (\delta_i X_i)^2 - \left(\frac{1}{n} \sum_{i=1}^n \delta_i X_i \right)^2$$

$$\begin{cases} X_i : \text{車両 } i \text{ の旅行時間} \\ \delta_i : 0 \text{ or } 1 \text{ の確率変数} \end{cases} \quad \sum_{i=1}^n \delta_i = n$$

ここで s_p^2 の期待値を考慮し数式を展開すると、

$$\begin{aligned} E(s_p^2) &= \frac{1}{n} \sum_{i=1}^N E(\delta_i^2) X_i^2 - \left(\frac{1}{n} \sum_{i=1}^N \sum_{j=1}^N E(\delta_i \delta_j) X_i X_j \right) \\ &= \frac{1}{N} \left(1 - \frac{1}{n} \right) \sum_{i=1}^N X_i^2 - \frac{1}{Nn} \left(\frac{n-1}{N-1} \right) \sum_{i \neq j}^N X_i X_j \\ &= \frac{N}{n} \left(\frac{n-1}{N-1} \right) \sigma^2 \end{aligned}$$

$$\therefore \sigma^2 = \frac{n}{N} \left(\frac{N-1}{n-1} \right) E(s_p^2)$$

上記のように、 $E(s_p^2)$ から σ^2 を推定することができ、これに補正項 $\frac{n}{N} \left(\frac{N-1}{n-1} \right)$ を加えることにより、 σ^2 の推定値を得ることができる。

断片的なプローブ軌跡の接合による区間旅行時間の期待値と分散の推定*

上杉友一**, 井料隆雅***, 小根山裕之****, 堀口良太*****, 桑原雅夫*****

本研究は、プローブ車両から得られるリンク旅行時間情報を用いて、ある対象区間の旅行時間を推定する際の、対象とするプローブ軌跡の選択方法と区間旅行時間の期待値・分散の推定精度の関係を検証するものである。すなわち、プローブ車両の混入台数と区間旅行時間の期待値・分散の推定精度の関係について、(1)「対象区間を完全に通過する車両の旅行時間情報だけを用いる場合」、(2)「(1)に加え、対象区間を部分的に通過する車両の旅行時間情報も考慮する場合」、(3)「(2)において、系統的な旅行時間の差異を含むリンク旅行時間を排除する場合」の3ケースを設定し、計算機実験を行うことで、各ケースの推定精度の考察を行うものである。

Estimation of Expectation and Variance for a Section Travel Time based on Fragmentary Probe Trajectories*

By Yuichi UESUGI**, Takamasa IRYO***, Hiroyuki ONEYAMA****, Ryota HORIGUCHI*****, Masao KUWAHARA*****

This paper discusses how to estimate the expectations and the variances of section travel times with sample vehicle trajectories, such as probes. Since the individual travel time of a section may differ from each other, an advanced travel information service (ATIS) should provide not only the accurate expectations of travel times but also the accurate variances. The accuracy of that information depends on the number of available probe trajectories. However, especially in low penetration ratio of probes, it is difficult to expect sufficient samples which traverse the whole section. Therefore, we should consider utilizing fragmentary trajectories which traverse some part of the section. In this paper, we describe the methodology of travel time estimation with fragmentary probe trajectories. Then the accuracy of the estimation is validated with computational experiment by changing sample ratio and sampling strategies.
