

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。

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。

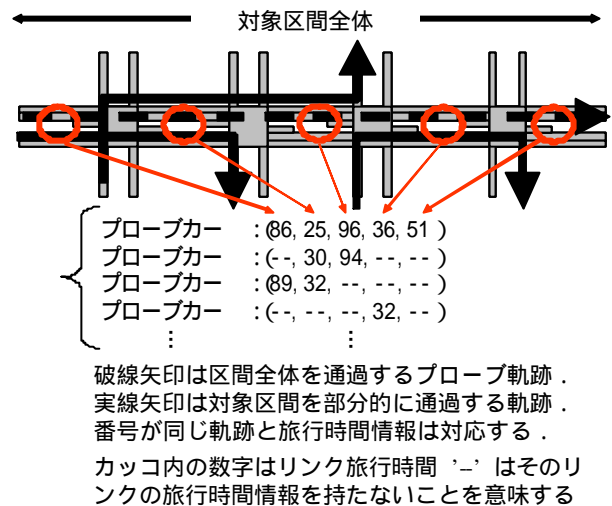


図1:完全又は断片的な旅行時間情報を持つプローブ軌跡

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。

2.1. リンク旅行時間から区間旅行時間を推定する方法

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。

$$m = \sum_{i=1}^l m_i \quad \dots\dots(1)$$

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

*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)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

m : 対象区間全体の旅行時間の期待値
 m_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

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.

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.

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

\hat{m}_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, $k=1,2,\dots,m$

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

$$a = \frac{N-1}{N} \cdot \frac{n_i}{n_i-1}, \quad b = \frac{N-1}{N} \cdot \frac{n_{ij}}{n_{ij}-1} : \text{補正項}$$

N : サンプリング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

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

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

\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, a, b は推定の際の統計的な補正項であり, そ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と, 式(1)における d_{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, d_{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, サンプリングさ

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の推定を行うものである.

式(3)についても同様の考え方で, サンプリング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.

2.2. ケース設定

以下では区間全体の旅行時間の期待値 m 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.

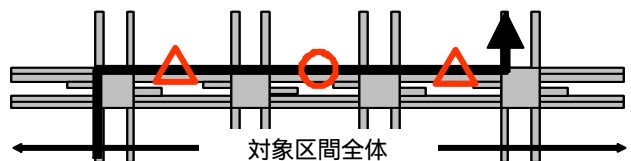


図2 : ケース2とケース3の違い
 (ケース3では の旅行時間だけを使用)
 (ケース2では と の両方を使用)

以上の3つのアプローチの概略を図3に示す.

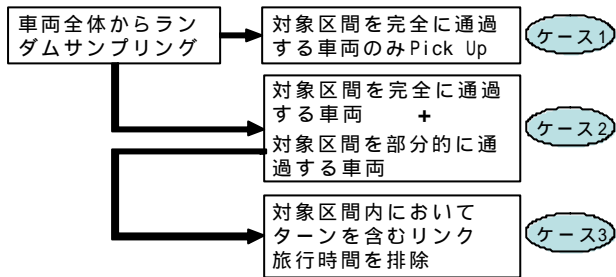


図3：3つのアプローチの概略

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により構成されるネットワーク上で非飽和交通流を再現した。

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。

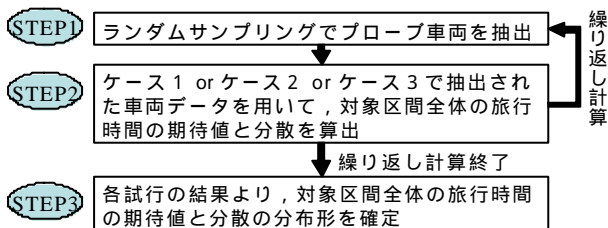


図4：計算機実験による検証手順

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 m 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のサンプル率 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、プローブ車両のサンプル率を変化させて、対象区間全体の旅行時間・分散についての計算機実験を行った結果を示し、その考察を行う。

表1：各手法を用いた場合の計算結果

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
ケース1	12196	205.0	100%	496.5		812.8	
	12074	203.0	99%	496.5	0.2	816.4	6.6
	6098	102.4	50%	496.5	2.0	813.7	64.5
	1219	20.4	10%	496.7	6.1	824.1	209.6
	609	10.4	5%	496.8	9.2	833.1	337.4
	243	4.1	2%	497.0	15.6	800.8	530.1
ケース2	12196	205.0	100%	482.0		1308.4	
	12074	202.9	99%	482.0	0.1	1308.7	5.5
	6098	102.5	50%	482.1	1.3	1310.4	56.0
	1219	20.4	10%	482.1	3.7	1306.1	168.4
	609	10.1	5%	482.5	5.5	1323.2	261.8
	243	4.1	2%	482.3	9.0	1293.5	422.3
ケース3	12196	205.0	100%	488.2		1031.8	
	12074	203.0	99%	488.2	0.1	1032.1	5.9
	6098	102.4	50%	488.3	1.4	1032.4	59.5
	1219	20.5	10%	488.1	4.3	1037.0	182.8
	609	10.4	5%	488.3	6.6	1041.0	294.1
	243	4.0	2%	487.9	10.9	1019.7	468.2

- (A)：サンプル車両台数
- (B)：(A)中の区間完全通過車両台数の試行毎平均
- (C)：サンプル率(%)
- (D)：区間全体の旅行時間の期待値 m (sec)
- (E)：(D)の標準偏差(s)
- (F)：区間全体の旅行時間の分散 s^2 (s²)
- (G)：(F)の標準偏差(s²)

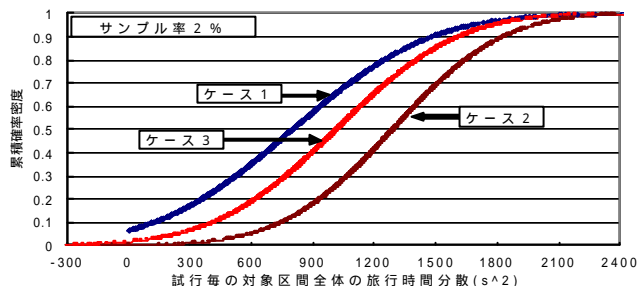


図5-1：サンプル率2%

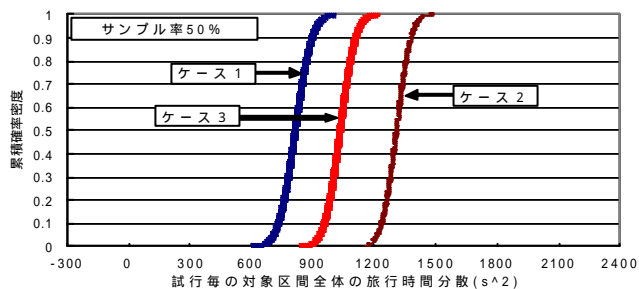


図5-2：サンプル率50%

図5：試行毎の対象区間全体の旅行時間分散値の累積分布表1には、試行を1000回行った際の、サンプル率の変化に対する区間全体の旅行時間の期待値・分散の変化を記した。

また、サンプル率が2%および50%の時について、繰り返し試行1000回時における、試行毎の対象区

間全体の旅行時間分散値の分布を示したものが図 5-1, 図 5-2 である。

まず, 区間旅行時間の期待値に注目すると, 表 1 より, 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, 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。

さらに, 図 5-1 と図 5-2 の両図を比較すると, 2% の分布は 50% のそれに比べ, 分布形がなだらかである。このことは, サンプル率が低いと, 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。

表 2 には, サンプル率が 0.5% の時と 10% の時の, 各ケースの推定の的中率を記してある。この場合的中とは, ケース 1 のサンプル率 100% 時の区間旅行時間の期待値(=496.5(s))を基準値とし, その値から $\pm 10(\text{sec})$, $\pm 20(\text{sec})$, $\pm 30(\text{sec})$ の各範囲に, 各ケースの試行毎の推定値が収まることであると定義する。

サンプル率が高い場合は, ケース 1 が最も的中率が高いといえる。表 2 において, ケース 2 の的中率が低いのは, ケース 2 の試行毎の旅行時間期待値の分布形が, 基準値(=496.5(s))から離れているためと考えられる(図 6 参照)。サンプル率が低下すると, 分布形はなだらかになるため, ケース 2 においては, サンプル率が小さいほうが基準範囲内に収まる試行が多くなる。

サンプル率が極端に低い時, ケース 1 においては, 絶対サンプルの少なさのため, 分散や共分散が計算不能な場合が生じ, 推定が不可能になるケースが増加する。この場合, 的中率においては, ケース 2, ケース 3 が上回ることが確認できる。

さらにサンプル率が低くなり, 対象区間を完全に通過する車両が存在しない場合は, ケース 1 は考

慮不可能となる。その場合には, ケース 2 および ケース 3 のみ考慮可能となる。

表 2: 各ケースにおける推定的中率(試行 1000 回中)

サンプル率0.5%時の的中率					サンプル率10%時の的中率				
範囲	± 10 (sec)	± 20 (sec)	± 30 (sec)	計算 不能	範囲	± 10 (sec)	± 20 (sec)	± 30 (sec)	計算 不能
ケース1	15.9%	34.5%	46.6%	37.2%	ケース1	88.5%	99.7%	100.0%	0.0%
ケース2	29.0%	52.9%	69.5%	12.8%	ケース2	11.0%	93.2%	100.0%	0.0%
ケース3	30.8%	55.7%	75.9%	4.6%	ケース3	64.8%	99.6%	100.0%	0.0%

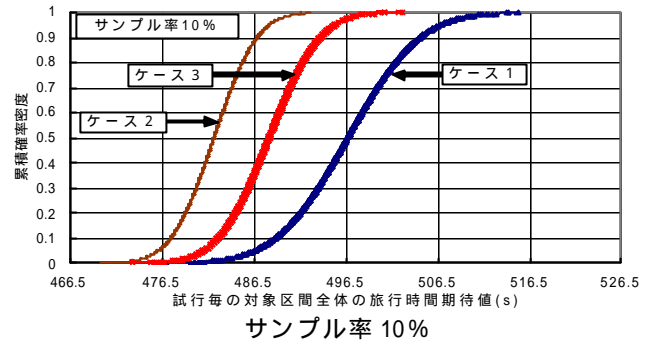


図 6: 試行毎の区間全体の旅行時間期待値の累積分布

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>

【付録】2章, 式(4), 式(5)中の補正項について

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。その際の分散 s^2 と, 分散の推定値 \hat{s}^2 は次のように書ける。

$$s^2 \equiv \frac{1}{N} \sum_{i=1}^N X_i^2 - \left(\frac{1}{N} \sum_{i=1}^N X_i \right)^2 \quad \left\{ \begin{array}{l} X_i: \text{車両 } i \text{ の旅行時間} \\ d_i: 0 \text{ or } 1 \text{ の確率変数} \end{array} \right.$$

$$\hat{s}^2 = \frac{1}{n} \sum_{i=1}^n (d_i X_i)^2 - \left(\frac{1}{n} \sum_{i=1}^n d_i X_i \right)^2 \quad \left\{ \begin{array}{l} \sum_{i=1}^n d_i = n \end{array} \right.$$

ここで推定値 \hat{s}^2 の期待値を考慮し, 数式を展開すると,

$$E(\hat{s}^2) = \frac{1}{n} \sum_{i=1}^n E(d_i^2) X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n E(d_i d_j) X_i X_j \right)$$

$$= \frac{1}{N} \left(1 - \frac{1}{n} \right) \sum_{i=1}^N X_i^2 - \frac{1}{Nn} \left(\frac{n-1}{N-1} \right) \sum_{i \neq j}^N X_i X_j$$

$$= \frac{N}{n} \left(\frac{n-1}{N-1} \right) s^2 \quad \therefore \frac{n}{N} \left(\frac{N-1}{n-1} \right) E(\hat{s}^2) = s^2$$

上記のように補正項を得る。