

# 交通流のモデルのパラメータの実験データからの推定

柴田章博

高エネルギー加速器研究機構 計算科学センター

## 概要

交通流のモデルのモデルパラメータを車の位置情報の時系列データから推定する方法を議論する。実際の交通流のデータは実験室での実験データと異なり必要な精度と時間間隔のものが必ずしも得られない場合が多い。このようなデータ欠損や非観測データが存在する場合においても、モデルの時間発展シミュレーションとを組み合わせたモデルパラメータの推定について考察する。

## Inference of the traffic model parameters from the time series data

Akihiro Shibata

Computing research Center, KEK

## Abstract

We study the statistical method to infer the model parameter from time series data of traffic flow. In the observation of realistic traffic flow, it is sometimes difficult to obtain the data with an expected accuracy or time interval. Here, we discuss the method to infer the model parameters even if some hidden variables or unobservable data are contained. The inference method which is combined with the model simulation is discussed.

## 1 はじめに

交通流の速度最適化モデル (optimal velocity (OV) model) は、高速道路の渋滞の生成と消滅を説明するモデルとして提唱された [1][2][3]。計算機上のシミュレーションでは、OV 模型のパラメータを適宜選択行って渋滞の形成を定性的に調べることが行われている。また、自動車をサーキット状に配置した渋滞形勢の実験が行われ、OV 模型で予言される渋滞の形成が行われている。[5] 実際の渋滞形成と交通流のモデルの対応を調べるためには、モデルのパラメータを実測のデータから決定しモデルの妥当性を調べることが重要である。図 1 に OV 模型によって生成された渋滞の例を示す。同様の渋滞形成の様子が、実験でも示されている。

本講演では、実験データを用いたモデル推定を行うため時系列データからのモデルパラメータの決定法について調べる。実際の実験では、運動方程式の時間発展を追跡することはできないので、より粗い時間間隔で

の観測に基づいて交通流の模型のパラメータを決定する方法について議論する。ここでは、実際のデータを用いた推定を行うための備的な研究とし、シミュレーションによって生成した時系列データを用いてパラメータ推定法の可能性について調べる。

## 2 パラメータ推定モデル

モデルパラメータの推定モデルとして、自動車の位置の時系列のデータ

$$\mathbf{Y} = \{y^t = (y_k)^t = (y_1^t, y_2^t, \dots, y_N^t)\} \quad (1)$$

ここで、 $k$  ( $= 1, 2, \dots, N$ ) は車両の番号を  $t$  ( $= 0, 1, 2, \dots, T$ ) は時間間隔  $t$  で観測した時系列の番号を表す。このとき次の速度最適化 (OV) モデルに従う時のモデルパラメータ  $(V_0, b, c, d)$  の推定を

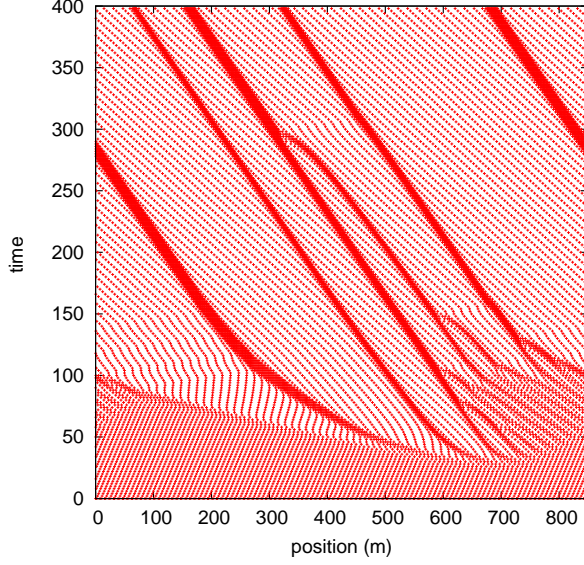


図 1: OV 模型における車両走行の軌跡。

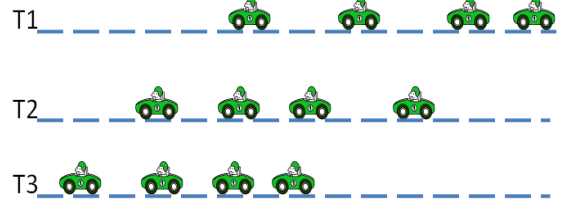


図 2: 車両走行のスナップショット

測定はすべての時間発展に対しては存在せず、 $t >$  であるとする。十分に細かな間隔で位置情報が得られれば、速度  $v_j^t$  が  $G_x(t, j) = 0$  によって計算でき、 $G_v(t, j)$  の最小自乗フィットによってパラメータを決定できる。ここでは、十分なデータが得られない場合おいてのパラメータ推定を考える。

考える。

$$\frac{d}{dt} x_i = v_i \quad (i = 1, 2, \dots, N) \quad (2)$$

$$\frac{d}{dt} v_i = \alpha_i [V(\Delta x_i; \alpha_i) - v_i] \quad (3)$$

$$OV(y; \alpha) = \max(v_0 \tanh(\frac{y-b}{d}) + c, 0) \quad (4)$$

ここで、 $x_i = x_{i+1} - x_i$  は前方の車両との車両間隔をあらわす。離散データの取り扱いを要するために、微分方程式をノイズを含む離散化された方程式（確率微分方程）のみに変形する。このとき蛙とび法による離散化を行い、 $\phi_j^{2t} = x_j^t$ ,  $\phi_j^{2t+1} = v_j^t$ , 離散化の時間間隔を  $\tau$  とすると方程式は次のようになる：

$$G_x(t, j) - \xi_j^t = 0, \quad G_x(t, j) - \eta_j^t = 0 \quad (5)$$

$$G_x(t, j) := \phi_j^{2t+2} - \phi_j^{2t} - \tau \phi_j^{2t+1} \quad (6)$$

$$G_v(t, j) := \phi_j^{2t+1} - \phi_j^{2t-1} - \tau a_j (OV(\Delta x_j^t) - \bar{v}_j^t) \\ = \alpha_j \phi_j^{2t+1} - \beta_j \phi_j^{2t-1} - \tau a_j OV(\Delta x_j^t; \alpha_j) \quad (7)$$

ここで、 $\bar{v}_j^t := (\phi_j^{2t+1} + \phi_j^{2t-1})/2$  とし、 $\alpha_j = 1 + a_j \tau / 2$ ,  $\beta_j = 1 - a_j \tau / 2$  とおいた。  $\xi_j^t$  及び  $\eta_j^t$  は Winner 型ノイズで、

$$\langle \xi_j^t \rangle = 0, \quad \langle \xi_j^t \xi_k^s \rangle = \sigma_x^2 \tau \delta^{ts} \delta_{jk}, \quad (8)$$

$$\langle \eta_j^t \rangle = 0, \quad \langle \eta_j^t \eta_k^s \rangle = \sigma_v^2 \tau \delta^{ts} \delta_{jk}, \quad (9)$$

を満たすとする。観測される位置情報  $y_j^s$  はノイズ  $\phi \sim N(0, \tau^2)$  を伴って  $\phi_j^{2t}$  観測されるとする。ただし、観

### 3 パラメータ推定

データ  $\mathbf{Y}$  からのパラメータ  $H = \{a_j, \alpha_j; j = 1, 2, \dots, N\}$  の推定は、ベイズ統計に従えば条件付き確率分布  $\Pr(H|\mathbf{Y})$  を求めることである。これは、尤度関数と  $L(\mathbf{Y}|H)$  と事前分布  $\Pr(H)$  を用いて次のように表せる。（ベイズの定理）

$$Pr(H|\mathbf{Y}) = \frac{L(\mathbf{Y}|H)Pr(H)}{P(\mathbf{Y})} = \frac{L(\mathbf{Y}|H)Pr(H)}{\int_H L(\mathbf{Y}|H)Pr(H)} \quad (10)$$

尤度関数は、 $Q^t = \{x_j^t, v_j^t\}$  を用いて、

$$L(\mathbf{Y}|H) = \int dQ L(\mathbf{Y}, Q|H) \quad (11)$$

と書くことができる。  $L(\mathbf{Y}|H)$  は、時間発展の変数  $Q$  によって尤度関数、  $L(\mathbf{Y}, Q|H)$  の周辺化を行ったものとして与えられるとした。  $L(\mathbf{Y}, Q|H)$  は、パラメータ推定を行う系全体を表すユニバーサル分布として解釈することができ、観測の独立性及び時間発展の因果律を用いて次のような条件付き確率の積として書くことができる。

$$L(\mathbf{Y}, Q|H) \\ = Pr(Q^1) Pr(Q^2|Q^1) \dots \prod_s Pr(Q^{s+1}|Q^s, H) \\ \times \prod_t Pr(Y^t|Q^{s(t)}, H) \quad (12)$$

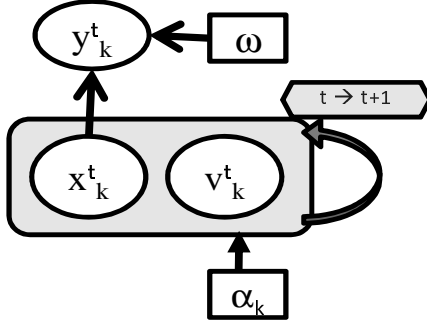


図 3: パラメータ推定のモデルのグラフィカル表現

$Pr(Y^t|Q^t, H)$  は観測過程であり、式 (9) のように分散  $\omega$  のガウス分布に従う：

$$\begin{aligned} Pr(Y^t|Q^t, H) &= \prod_j Pr(y_j^t|x_j^t, H) \\ &= \frac{1}{2\sqrt{\pi\omega}} \exp\left(-\frac{1}{2\omega^2} \sum (y_j^t - x_j^t)^2\right) \end{aligned} \quad (13)$$

ここで、 $s(t)$  は、観測データの時刻  $t$  に対応する方程式の時刻  $s$  を指す。 $Pr(H)$  は位置データの観測に関する事前分布を与える。 $Pr(Q^{t+1}|Q^t, H)$  は、モデル  $H$  における時間発展であり、確立微分方程式 (5) に従う。 $\eta$  は正規分布に従うから、次のように与えられる。

$$\begin{aligned} Pr(Q^{t+1}|Q^t, H) \\ \propto \exp\left(-\sum \frac{G_x(t, j)^2}{2\sigma_x^2\tau} - \sum \frac{G_v(t, j)^2}{2\sigma_v^2\tau}\right) \end{aligned} \quad (14)$$

$Pr(Q^1)$  は、運動方程式の初期条件を与える事前分布である。ここでは、事前分布として  $Q^1$  及び  $H$  は、無事前情報の仮定  $Pr(Q^1) \propto 1$ ,  $Pr(H) \propto 1$  と においては Jeffy のクライテリア  $Pr(H) \propto \omega^{-N^2+1}$  をとる。したがって、パラメータ推定の確率モデル (事後分布) は次のように与えられる。

$$Pr(H, Q|Y) = \exp(-M(Q, Y; H))Pr(H), \quad (15)$$

$$\begin{aligned} M(Q, Y; H) &= \sum_{s,j} \left( \frac{G_x(t, j)^2}{2\sigma_x^2\tau} + \frac{G_v(t, j)^2}{2\sigma_v^2\tau} \right) \\ &+ \frac{1}{2\omega^2} \sum_{j,s} \theta_s (y_j^{t(s)} - \phi_j^{2s})^2 \end{aligned} \quad (16)$$

ここで、 $s$  は観測データが存在する時刻において  $s = 1$  であり、その他はゼロをとる。

図 (3) は、パラメータ推定表現のグラフィカル表現を表す。

## 4 パラメータ決定アルゴリズム

$Pr(H|Y)$  を最大化することは、周辺化事後分布 (11) の積分を実行する必要がある。しかしながらこの積分の実行は多次元積分の実行を伴い計算の実効性に欠ける。代わりに、 $Pr(H, Q|Y)$  非観測データ、 $Q$  を含む事後分布を最大化を行う。 $Pr(H, Q|Y)$  は運動方程式解の周りに鋭いピークを持つ。非観測量を含む尤度関数に対する最尤法にアルゴリズムとして EM アルゴリズムがある。この EM アルゴリズムを用いることで  $Pr(H, Q|Y)$  の最大事後分布を与えるパラメータを決定できる。EM アルゴリズムは、反復法によってパラメータを推定する。モデルメータ  $H^{(k)}$  とデータ  $Y$  が与えられた時の変数  $Q$  がの出現確率はベイズ定理より

$$Pr(Q|H, Y) = \frac{Pr(H, Q; Y)}{\int dQ' Pr(H, Q'; Y)} \quad (17)$$

と与えられる。ここで、 $Pr(Q|H, Y)$  で重みを付けて周辺化した”尤度”

$$\begin{aligned} R(H'; H^{(k)}, Y) \\ := \int dQ \log (Pr(H', Q; Y)) \\ \times \frac{Pr(H^{(k)}, Q; Y)}{\int dQ' Pr(H^{(k)}, Q'; Y)} \end{aligned} \quad (18)$$

を定義すると EM アルゴリズムは次のように構成される。(詳しくは、たとえば 文献 [6] を参照のこと)

アルゴリズム 1 *begin loop*

*E-step:*  $R(H'; H^{(k)}, Y)$  を計算する。

*M-step:*  $H^{(k+1)} = \arg \max_{H'} R(H'; H^{(k)}, Y)$

*until*  $\|H^{(k+1)} - H^{(k)}\| \leq \epsilon$

*return*  $H = H^{(k+1)}$

EM アルゴリズムの実行には、式 (18) の評価のため式 (17) 重みのついた多次元積分が必要となる。この積分の実行をモンテカルロ法によって実行する。サンプルポイント  $\{x\}$  が確率分布関数  $Pr(x) = \rho(x)dv$  で生成されるとき、多重積分は次のようにアンサンブルで計算することができる。

$$\int f(x)dv = \int \frac{f(x)}{\rho(x)}\rho(x)dx \quad (19)$$

$$= \int g(x)\rho(x)dx = \langle g(x) \rangle_\rho, \quad (20)$$

$$\langle g \rangle_\rho := \frac{1}{N} \sum_{\{x\}} g(x). \quad (21)$$

ここで、 $N$  は サンプルポイントの数で、期待値の統計的エラーは、

$$Error = \frac{1}{\sqrt{N}} \sqrt{\langle g(x)^2 \rangle_\rho - \langle g(x) \rangle_\rho^2} \quad (22)$$

であたえられる。本モデルにおいては、

$$\begin{aligned} g(Q) &= \log(\Pr(H', Q; \mathbf{Y})) \\ &= -M(Q, \mathbf{Y}; H') \end{aligned} \quad (23)$$

$$\begin{aligned} \rho(Q) &= \frac{\Pr(H, Q; \mathbf{Y})}{\int dQ' \Pr(H, Q'; \mathbf{Y})} \\ &= \frac{\exp(-M(Q, \mathbf{Y}; H'))}{\int dQ' \exp(M(Q, \mathbf{Y}; H^k))} \end{aligned} \quad (24)$$

で与えられる。重点サンプル法で成分を実行するためには、 $\rho(Q)$  でサンプルポイントを生成しなければならぬ。式(24)は、まさに統計力学における分配関数の形しており、統計力学のモンテカル・シミュレーションの技法が有用である。

## 5 シミュレーション

モデルパラメータ推定法の実効性を調べるためにここではシミュレーションのデータを用いる。実際の実験データを用いた場合には、推定を行うモデルとデータの適合性の問題があり、また時系列のデータのサンプリングの間隔のコントロールが難しい。シミュレーションで生成したデータを用いることで、これらの問題を回避しメソッド実用性を見ることが出来る。図1は、OVモデルのシミュレーションで得られた渋滞の様子をあらわしている。車の軌跡は、シミュレーションのタイムステップのオーダーでサンプリングしたデータをプロットした。

サンプリング間隔が十分細くとれる”理想的な”場合、すなわち、サンプリングの間隔が微分方程式の時間発展のステップと等しい場合においては  $v_j^t = (x_j^{t+1} - x_j^t)/\tau \cong (y_j^{t+1} - y_j^t)/\tau$  と速度が決定される。また、 $\sum_{j,t} G_v(t, j)^2$  の最小化条件、 $\chi^2$  フィットによって決定できる。車両の位置、車間、速度が与えられるため、モデルパラメータのフィットは各車両ごとに独立に行うことができる。

一方、十分細かな時間間隔でデータを得られない場合には、EMアルゴリズムを実行する必要があるが、式(24)の重みによる重点サンプルに積分は、統計力学などで用いられるモンテカルロシミュレーションの技法を導入する。変数  $Q_j^t$  のメロポーリス法によるローカルなアップデートによって、車両全体の作るトラジェ

クトリー  $\{Q_j^t\}$  を生成し、モンテカルロ積分によって式(18)を評価することができる。

## 6 まとめ

車両位置の時系列データに基づく交通流モデルのパラメータ推定の統計モデルを構成した。観測データが十分存在しない場合についてもモデルに基づいた確率論的な推定を行えるモデルを検討した。推定法の有効性を調べるために、OVモデルのシミュレーションによって生成したデータを用いて、データのサンプリング間隔をかえてシミュレーションのモデルパラメータの推定を行った。車両速度を求めることができる十分細かな時間間隔でデータを採取できる場合は、通常  $\chi^2$  フィットに帰着しパラメータ推定を行うことができた。サンプリングの間隔が開いた場合については、現在解析を進めており、シンポジウム講演のときに報告をする。

## 参考文献

- [1] M.Bando, K.Hasebe, A.Nakayama, A.Shibata and Y.Sugiyama, Japan. J. of Industrial and Applied mathematics. Vol 11 No2 203-233 (1994)
- [2] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, Phys.Rev. E 51, 1035-1042 (1995)
- [3] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, J. Phys. I France 5 1389-1399 (1995)
- [4] A. Shibata, Proceedings of International workshop Traffic and Granular Flow '01 pp.91-96, Springer ISBN 3-540-40225-1
- [5] Minoru Fukui, Katsuya Hasebe, Macoto Kikuchi, Akihiro Nakayama, Katsuhiko Nishinari, Yuki Sugiyama, Shin-ichi Tadaki, Satoshi Yukawa, presented at the conference "Traffic and Granular Flow" 20-22, June 2007, Paris-Sud University, Orsay
- [6] R.O. Duda, P.E. Hart and D.G. Stork, Pattern Classification 2nd Edit., A Wiley-Interscience Publication.
- [7] Reference for the Monte Carlo integral, for example, Numerical recipes inC/C++