第Ⅳ部門　　プローブ観測を活用した検知器データのベイズ補完

神戸大学　学生員　○杉田　薫子
神戸大学大学院　正会員　長江　剛志
神戸大学大学院　正会員　朝倉　康夫

１．はじめに

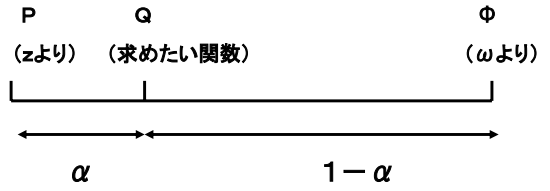
本研究では，検知器データとプローブデータを効果的に組み合わせ，高速道路上の速度分布を推定する新たな方法論を提案する．現在，検知器により大量の速度データが取得されている．しかし，検知器は走行車線，追越車線すべてに等間隔で設置されているわけではない．これに対して，近年注目を集めているプローブは，経済的に安価であり，柔軟な情報収集を行なうことができる．そこで，検知器データとプローブデータを用いて不足しているデータを補完するモデルを構築するという発想は極めて自然である．

このモデルは検知器データの事前速度分布とプローブデータの事前速度分布の「内分点」として，より精度の高い事後速度分布を推計するものである．本研究では，この直感的手法が，距離の指標として相対エントロピーを用いることで，従来の伝統的ベイズ推計モデルと等価な数理構造を持つことを明らかにする．

２．ベイズ補完モデル

２—１　モデルの枠組み

追越車線の検知器データをｚ，走行車線のプローブデータをωとする．この２種類のデータを用いて，走行車線の速度推定を行なう．ｚ，ωそれぞれについて確率密度関数P,Φを描く．ｚは検知器より大量に取得できるので，発生回数を数え上げることによってPを描く．ωは非常にデータ数が少ないと考えられるので正規分布を仮定し,パラメータを特定化することによってΦを描く．求めるべき走行車線の推定速度の確率密度関数をQとする．この確率密度関数Qは，前述の２つの確率密度関数P,Φの間に位置すると考える．どちらの確率密度関数に重みをおくかを決めるパラメータαを，使用データとプローブデータ数による所与のものとし，２つの確率密度関数より新しい確率密度関数を求めるイメージを図．１に示す．



図．１

Qの算出方法としては，プローブデータ数が少なければ検知器データより算出されたPに重みをおき，プローブデータ数が多くなればなるほどプローブデータより算出されたΦに重みをおいた上で，PとQ，ΦとQのグラフ間距離の和の最小化を行なうというものである．

２－２　更新プロセスの数学的導出

確率密度関数P,ΦをＫ個に分割し離散的に捉える．ここで，Ｋ個に分割したP,Φのk番目の長方形の面積すなわち確率をそれぞれ，$p(k),\varphi(k)$とする．また，Ｋ個に分割した求めるべき確率密度関数Qのk番目の長方形の面積，すなわち確率を$q(k)$とする．２つの関数QからPへの距離を$H(Q,P)$，QからΦへの距離を$H(Q,\Phi)$とおく．相対エントロピーの定義を用いると次のように書ける．

$$H(Q,P)=\sum_k q(k)\ln\left[\frac{q(k)}{p(k)}\right]$$

$$H(Q,\Phi)=\sum_k q(k)\ln\left[\frac{q(k)}{\varphi(k)}\right]$$

この２つの各グラフ間の距離のどちらにどれだけの重みをおくかを考慮した上で，グラフ間の距離を最小化する．この重みを表すのがパラメータαである．以下に最小化問題の定式化を示す．

$$\underset{\{q(k)\}}{Min.}\quad (1-\alpha)\sum_k q(k)\ln\frac{q(k)}{p(k)}+\alpha\sum_k q(k)\ln\frac{q(k)}{\varphi(k)}$$

$$s.t.\quad \Sigma q(k)=1,\quad q(k)>0$$

Kaoruko SUGITA, Takeshi NAGAE, Yasuo ASAKURA

この最小化問題を解くと下のようになる.

$$q(k)=\frac{r(k|\varphi)\cdot p(k)}{\Sigma\{r(k|\varphi)\cdot p(k)\}}$$

ただし, $r(k|\varphi)=p(k)^{\alpha}\cdot\varphi(k)^{-\alpha}$

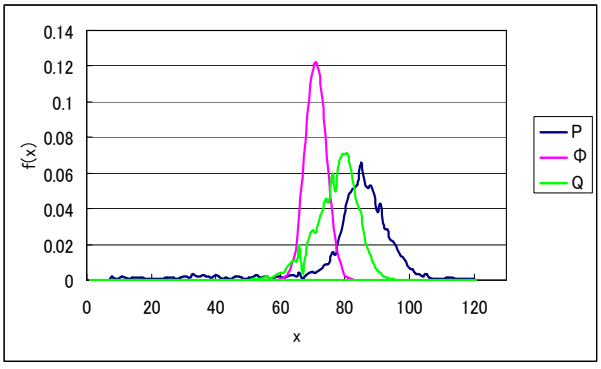
これは，尤度$r(k\mid\varphi)$、事前確率$p(k)$とするベイズ更新理論の式に他ならない.

この式にしたがって，すべてのkについて$q(k)$を算出する．求められた確率$q(k)$を，分割した幅に応じてグラフ化すれば，確率密度関数を描くことができ，連続的な走行速度の発生確率を求めることができる．このモデルは，プローブデータ数に関わらず，１つの手法でシームレスに対応できるものである.

３．実証分析

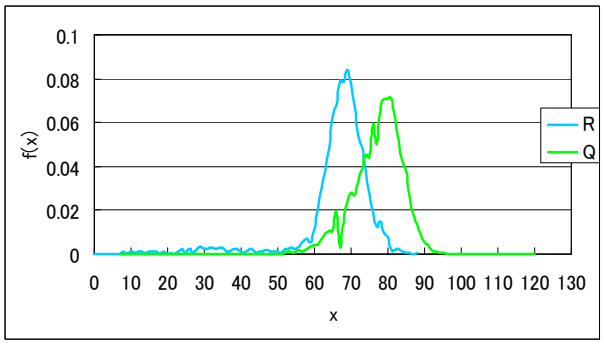
実証分析では，モデルの適合性を検証するため，追越車線，走行車線共に検知器が存在する阪神高速道路11号池田線上り5.1ｋｐに着目した．追越車線の検知器データを「検知器データ」とし，走行車線の大量の検知器データから任意の数のデータを無作為に取り出したものを「プローブデータ」とした.

「検知器データ」より描いた確率密度関数P，「プローブデータ」より正規分布を仮定して描いた確率密度関数Φ，求められた確率密度関数Qを図．２に示す.



図．２

このモデルを適用することによって，走行車線の推定速度についての確率密度関数のグラフが描けた．分析の結果と実データとの適合性を検証するため，走行車線の検知器データについての確率密度関数のグラフRを描き，図．３にQと共に示す.



図．３

このグラフQを読み取ることによって，走行車線の速度データの発生確率が求められる.

Qは，外形としては真実のグラフRと似通ったものが求められた．しかし，１番起きやすい速度が真実の値より，10km/h 程度ずれて算出された．原因としては，Φを正規分布で近似したことが考えられる．また，Qは x=50 以下，x=100 以上の部分では，ほぼ０になっている．これは，QはPとΦの掛け合わせによるので，どちらかが０をとった場合，Qも０になってしまうことが原因と考えられる.

４．おわりに

このモデルは，特定の１地点のみならず，より広範囲における検知器データを使用するように拡張できる可能性を秘めている．n 次元正規分布を用いて，走行車線，追越車線についての確率密度関数をそれぞれ描けば，ｎヶ所の地点における速度データを用いて速度を推定することができる．拡張することによるメリットとしては，データ数が大幅に増加するので，より正確な速度推定が可能になること，また連続した地点の速度データを用いるので，連続的な交通の流れも考慮できることが挙げられる.

本研究のモデルは概念的には，「ある部分の情報を用い，他の似通っているであろう部分の情報を補完する」といえる．本手法の基本姿勢は非常に普遍的であり，検知器データとプローブデータの関係にとどまらず，様々な分野に活用できるものであると考える.

【参考文献】

１）薩摩順吉：確率・統計，岩波書店，1989
２）小島寛之：サイバー経済学，集英社，2001
３）都築卓司：なっとくする統計力学，講談社サイエンティフィク，1998