

卒論テーマ候補：混雑現象の構造推定

池上 慧

2017 年 11 月 9 日

1 目的

Viauroux (2007) や柳沼・福田 (2007,8)、松村他などが交通手段の選択と通勤に利用する電車の出発時間選択においてプレイヤー同士の影響を考慮したゲームのモデルを用いたパラメータの推定を行っている。またその結果においては混雑による不効用を示すパラメータは有意に負の値を取っている。主に用いられていた推定手法は Agurregabiria (2004) の PML であり、Kasahara and Shimotsu(2012?) にのっとって推定結果が収束するようにしている。

上記の研究ではプレイヤーについての情報が得られている時にプレイヤーごとに選択確率を計算することでベイジアンナッシュ均衡を計算し、推定に用いている。しかし、松村他で用いられたデータはサンプルサイズが小さく、センサスデータであり、また個人の通勤時間などについては別種の統計データから推計したものを利用している。ここではこのような意思決定を行う個人に関するデータを必要とせずに混雑現象を推定できるかを考える。

例えば、松村他で扱われた満員電車に選択モデルを考えると、個人ごとにどのような人がいつどの電車に乗ったかというデータを得るのは非現実的である。そのようなデータの代わりに何時の電車に何人が乗ったという現実的に容易に得られるデータを用いて推定できるモデルを作ることは有用であると考ええる。

現実存在する混雑現象としては、上で述べた満員電車の他にも、入試の受験料を収入源とする私立大学群による入試日の設定や、テーマパークへの来場日の決定、コンピュータ上の分散処理などが挙げられる。このようなテーマで実際にモデルを推定する。

2 モデル

例としてわかりやすいので満員電車の事例でモデルを記す。 M 駅の T 日分の乗客数データが得られているとする。ここで $m \in \{1 \cdots M\}$ で 1 つの駅を示すとして、そこで乗車する総人数を N_m で記す。各駅は電車を 2 本持っていて、プレイヤーこの 2 本のうちより効用の高い方を選んで乗車する。駅 m を利用する個人 i の電車 j に対しての効用は以下で書けるとする。

$$X_i' \beta + Z_j' \gamma + \eta_j^i + \alpha B_j^m + \epsilon_j^i$$

ここで、 X_i はプレイヤーの属性を含むベクトル、 Z_j は電車 j の属性を含むベクトル、 η_j^i はプレイヤーごとに持つ電車 j に対する観測不可能な選好 (public information)、 B_j^m は駅 m における電車 j への実現した乗車人数、 ϵ_j^i はプレイヤーが私的情報として持つ電車 j への選好をそれぞれ表す。

離散選択の定式化を利用するため、 ϵ_j^i は i, j それぞれに対して独立に同じ第 1 種極値分布に従うとする。この rational expectation の仮定の下で、各駅に置いて以下の方程式を満たす形で毎日均衡としての乗車人数の期待値 ($E[b_{i,m}^{N_m}]$) が決定する。ただし、 $d_i^t = \eta_2^i - \eta_1^i$ である。

$$E[b_{i,m}^{N_m}] = \frac{1}{N_m} \sum_{i=1}^{N_m} \frac{1}{1 + \exp \left((z_2^m - z_1^m)' \gamma + d_{i,m}^t + \alpha N \left(1 - 2E[b_{i,m}^{N_m}] \right) \right)}$$

ここで、 $\widehat{E[b_m^{N_m}]} = \frac{1}{T} \sum_{t=1}^T E[b_{i,m}^{N_m}]$ と書く。 $T \rightarrow \infty$ におけるこの収束先について考える。駅 m を利用する個人 i が

電車 1 を利用する確率を離散選択の定式化で求めたものを $p_{1,m}^i$ で記す。この時以下が成立する。

$$p_{1,m}^i > \frac{1}{2} \Leftrightarrow \frac{1}{1 + \exp\left((z_2^m - z_1^m)' \gamma + d_i + \alpha N \left(1 - 2E[b_m^{N_m}]\right)\right)} > \frac{1}{2}$$

$$\Leftrightarrow d_i < \alpha N_m (2E[b_m^{N_m}] - 1) - (z_2^m - z_1^m)' \gamma$$

ただし、 $E[b_m^{N_m}]$ は heterogeneous な個人についても平均的な見た時にモデルの均衡として求まる乗車割合である。 d_i の従う分布の CDF を $F(\cdot)$ とすると、モデルの均衡は以下の方程式を満たすことがわかる。

$$E[b_m^{N_m}] = F\left(\alpha N_m (2E[b_m^{N_m}] - 1) - (z_2^m - z_1^m)' \gamma\right)$$

ここから即座に以下を得る。

$$E[b_m^{N_m}] = \frac{1}{2} + \frac{(z_2^m - z_1^m)' \gamma + F^{-1}(E[b_m^{N_m}])}{2\alpha N_m}$$

これより、分布 $F(\cdot)$ がサポートが有限で、 $\{z_j\}$ が有限のサポートを持ちかつパラメータ γ も有限であるなら、 $N \rightarrow \infty$ で $E[b_m^{N_m}] \rightarrow \frac{1}{2}$ であることがわかる。

一方で、最尤推定量のアナロジーから適切な条件の下で以下は成立しそう（要証明）

Claim

d_i^t が同じ分布 $F(\cdot)$ から生成されている時、 $\widehat{E[b_m^{N_m}]^t} \xrightarrow{P} E[b_m^{N_m}]$

この主張が正しければ、 T, N_m が十分に大きい時に以下が成立する。

$$\frac{1}{2} = \frac{1}{T} \sum_{t=1}^T \frac{1}{N_m} \sum_{i=1}^{N_m} \frac{1}{1 + \exp\left((z_2^m - z_1^m)' \gamma + d_{i,m}^t + \alpha N \left(1 - 2E[b_{t,m}^{N_m}]\right)\right)}$$

ここで $d_{i,m}^t$ の分布をベータ分布 ($Beta(a, b)$) であるとする。ここから発生させた乱数を用いて先の条件が全駅で達成されるように目的関数を最適化することによりパラメータの推定値を求めることを考える。具体的には $Beta(2, 2)$ を proposal distribution とする importance sampling を用いて先の右辺を構成する。ここで乱数の発生回数を R とする。 R を増やすことは十分大きな N_m の下では結果的に N_m をより大きくするような働きをするので先の条件が成立するために必要な環境を整えることになる。

以上より駅 m において満たすべき式は、 $S_m = N_m R$ として、以下のように書ける。

$$\delta_m \equiv \frac{1}{2} - \frac{1}{T} \sum_{t=1}^T \frac{1}{S_m} \sum_{s=1}^{S_m} \frac{1}{1 + \exp\left((z_2^m - z_1^m)' \gamma + x_{s,m}^t + \alpha N \left(1 - 2E[b_{t,m}^{N_m}]\right)\right)} \frac{(x_{s,m}^t + \frac{1}{2})^{a-2} (\frac{1}{2} - x_{s,m}^t)^{b-2}}{6B(a, b)} = 0$$

これが全駅について成り立つようにパラメータを推定するので、以下の最適問題の解がパラメータの推定値である。

$$\min_{\alpha, \gamma, a, b} \sum_{m=1}^M \delta_m^2$$

これは KNITRO などの solver で扱える問題である。

3 シミュレーション

Julia でシミュレーションデータを作成した。最適化は solver が起動できず未達成。正しいパラメータを入れれば目的関数が極めて小さな値となることは確認できた。また、データ作成の過程で先の claim も成り立ちそうなことを確認。

4 今後

- K 選択肢に拡張
- 乗客数データを用いて実データから推定。時刻表変更の反実仮想を行う。