

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 研究目的

Viaurox (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のはサンプルサイズの小さい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個人のデータによら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得られるデータのみ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（本当？）。

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

2.1 基本のモデル

電車の候補が全部で M 本とし、 $j \in M$ で各電車を指すとする。プレイヤーは N 人で i をインデックスとする。プレイヤー i が電車 j に乗車することの効用 (V_j^i) は、その電車の発車時刻 (t_j) に紐付いた i ごとに異なる価値 (u_j^i) と、その電車に乗車するプレイヤーの数 (N_j) に依存して決定する。ただし $\{u_j^i\}$ は自身の持つ各電車への価値は自分で認識できるが、自分以外のプレイヤーが持つ各電車への価値は認識できないとする。ここでそれぞれの効用への寄与は分離できるとして、

$$V_j^i = u_j^i + g(N_j)$$

で書くことができるとする。

しかし、実際に乗車する電車を決定する際には、実現する乗車人数 N_j に基づいて意思決定することはできない。なぜならまだその人数は実現していないからである。そのため、 N_j についての期待値に基づいて意思決定を下すことになる。ここで、プレイヤー同士にも、経済学者にも観測できない攪乱項 ϵ_j^i が電車とプレイヤーについて独立に第 1 種極値分布にしたがうと仮定し、*additive* に効用に影響するとする。すなわち、

$$h_j^i = V_j^i + \epsilon_j^i = u_j^i + g(N_j) + \epsilon_j^i$$

でプレイヤー i の持つ電車 j の効用をかけるとし、 N_j について期待値をとった形で

$$h_j^{i*} = V_j^{i*} + \epsilon_j^i = u_j^i + E[g(N_j)] + \epsilon_j^i$$

と書くとする。この時、 $g(N_k) = \alpha N_k$ として特定化すると、プレイヤー i が電車 j を選ぶ確率は以下のようにかける。

$$p_j^i = \Pr(i \text{ chooses } j) = P(h_j^{i*} \geq h_{j'}^{i*} \text{ for all } j' \in M) = \frac{\exp(V_j^{i*})}{\sum_k \exp(V_k^{i*})} = \frac{\exp(u_j^i + \alpha E[N_j])}{\sum_k \exp(u_k^i + \alpha E[N_k])}$$

2.2 N 人 2 物ゲーム

N 人 2 電車の状況を考える。Brock and Durlauf (2001) で記述されているようなモデルを考えていく。 p_k^i で、プレイヤー i が電車 k に対して持つ選択傾向を表すとする。第 1 種極値分布に従う攪乱項を用いてそれぞれの電車に対する選択傾向を以下のように書ける。

$$\begin{cases} p_1^1 &= \frac{\exp(u_1^1 + \alpha E^*)}{\exp(u_1^1 + \alpha E^*) + \exp(u_2^1 + \alpha(N - E^*))} \\ \vdots & \\ p_1^N &= \frac{\exp(u_1^N + \alpha E^*)}{\exp(u_1^N + \alpha E^*) + \exp(u_2^N + \alpha(N - E^*))} \end{cases}$$

この時 E^* は以下で述べていく意味での均衡における電車 1 に乗車した人の数であるとする。この選択傾向が $\frac{1}{2}$ を上回れば電車 1 を選択し、下回れば電車 2 を選択するとして行動を記述する。この時、プレイヤー $i \in \{1, \dots, N\}$ について、電車 2 を選ぶということは

$$\begin{aligned} p_1^i < \frac{1}{2} &\Leftrightarrow \frac{\exp(u_1^i + \alpha E^*)}{\exp(u_1^i + \alpha E^*) + \exp(u_2^i + \alpha(N - E^*))} < \frac{1}{2} \\ &\Leftrightarrow u_2^i - u_1^i > 2\alpha E^* - \alpha N \\ &\Leftrightarrow d_i > 2\alpha E^* - \alpha N \end{aligned}$$

である。ただし $d_i = u_2^i - u_1^i$ としている。ここで $\forall i \ d_i \sim N(0, \sigma^2)$ とおくと、 $\Pr(d_i > \alpha(2E^* - N)) = 1 - \Phi\left(\frac{\alpha}{\sigma}(2E^* - N)\right)$ で書くことができる。以上より均衡における電車 1 に乗る人の期待値 E^* は、

$$N - E^* = N(1 - \Phi\left(\frac{\alpha}{\sigma}(2E^* - N)\right)) \Leftrightarrow E^* = N\Phi\left(\frac{\alpha}{\sigma}(2E^* - N)\right)$$

をを満たすことがわかる。つまり $N\Phi\left(\frac{\alpha}{\sigma}(2E - N)\right)$ であり、ここで $e^* = \frac{E^*}{N}$ すると $e^* = \Phi\left(\frac{\alpha}{\sigma}N(2e^* - 1)\right)$ を均衡の e^* は満たすことになる。

2.3 推定

2.3.1 正規性の仮定の下で考える

今、 $v = e - \Phi\left(\frac{\alpha}{\sigma}N(2e - 1)\right)$ とおくと、 $\alpha < 0$ であるという想定の下では、

$$\frac{\partial v}{\partial e} = 1 - \frac{2\alpha N}{\sigma} \phi\left(\frac{\alpha}{\sigma}N(2e - 1)\right) > 1$$

を得るので、 v は e についての単調増加関数である。従ってこのモデルの均衡はただ一つであり、 $e = \frac{1}{2}$ において先の等式が成立することから、 $e^* = \frac{1}{2}$ が唯一の均衡である。この時 α がどのような値でも同じ均衡がモデルから得られるのでパラメータの識別は行えない。これは random component である $\{u_j^i\}$ の平均が 0 である事、すなわちプレイヤーが平均的には二つの選択肢について等しく価値を置いている事がもたらす問題である。

次に $d_i \sim N(\mu, \sigma^2)$ で $\mu \neq 0$ であるケースを考える。上と同様の計算により、

$$\begin{aligned} e^* &= \Phi\left(\frac{\alpha N(2e^* - 1) - \mu}{\sigma}\right) \\ \Leftrightarrow \alpha &= \frac{\sigma \Phi^{-1}(e^*) + \mu}{N(2e^* - 1)} \end{aligned}$$

を得る。ここで $\alpha < 0$ であるために必要な条件として

- $e^* < \frac{1}{2}$ ならば $\mu > 0$
- $e^* > \frac{1}{2}$ ならば $\mu < 0$

が必要である事がわかる。これはすなわち、混雑現象の推定においては平均的に見てどちらかの選択肢の方が好まれているような状況下において、それでも混雑を避けて異なる選択肢を選ぶ人の方が多くなる時にパラメータの識別が可能になる事を意味する。

2.3.2 一般的な分布で考える

先のモデルの解である $e \in [0, 1]$ の挙動を考える。 $e \sim g(\cdot)$ として、累積分布関数を $G(\cdot)$ で表記することにする。 $d_i \sim f(\cdot)$ で、その累積分布関数を $F(\cdot)$ で書くとする。この時モデルの均衡である e はその期待値が $E[e] = F(\alpha N(2E[e] - 1))$ を満たす。これを用いて $E[e]$ は以下のようにして求められる。

$$\begin{aligned} E[e] &= \int_0^1 (1 - G(z)) dz = F(\alpha N(2E[e] - 1)) \\ \Leftrightarrow \alpha N(2E[e] - 1) &= F^{-1} \left(1 - \int_0^1 G(z) dz \right) \\ \Leftrightarrow E[e] &= \frac{1}{2} + \frac{F^{-1} \left(1 - \int_0^1 G(z) dz \right)}{2\alpha N} \end{aligned}$$

d_i のサポートが有限区間であると仮定すると、 $F^{-1} \left(1 - \int_0^1 G(z) dz \right)$ は有限なので、 N を無限大にした時に $E[e] = \frac{1}{2}$ となる。

次に e の分布を考える。

$$\begin{aligned} Pr(e < z) &= Pr \left(1 - \frac{\# \{d_i > \alpha N(2e - 1)\}}{N} < z \right) \\ &= Pr \left(1 - z < \frac{\# \{d_i > \alpha N(2e - 1)\}}{N} \right) \end{aligned}$$

ここで $\# \{d_i > \alpha N(2e - 1)\} \sim Bin(N, 1 - F(\alpha N(2e - 1)))$ であることより、上式で N についての極限を取り、

$$Pr \left(1 - z < 1 - \lim_{N \rightarrow \infty} F(\alpha N(2e - 1)) \right)$$

適切な条件の下で極限と積分は順序を交換できること、そして $\alpha < 0$ であることより

$$Pr \left(F \left(\lim_{N \rightarrow \infty} \alpha N(2e - 1) \right) < z \right) = Pr \left(\lim_{N \rightarrow \infty} \alpha N(2e - 1) > \frac{F^{-1}(z)}{\alpha} \right)$$

を得る。以上より N を大きくした時、 $G(z) = Pr \left(\lim_{N \rightarrow \infty} \alpha N(2e - 1) > \frac{F^{-1}(z)}{\alpha} \right)$ であることがわかる。

ここで $x = \lim_{N \rightarrow \infty} \alpha N(2e - 1)$ として、その累積密度関数を $H(\cdot)$ 、密度関数を $h(\cdot)$ と書くことにする。この時、

$$g(z) = \frac{dG}{dz} = \frac{d}{dz} \left(1 - H \left(\frac{F^{-1}(z)}{\alpha} \right) \right) = -\frac{1}{\alpha f(F^{-1}(z))} h \left(\frac{F^{-1}(z)}{\alpha} \right)$$

を得る。

x の分布を求める。先の議論より $e \xrightarrow{P} \frac{1}{2}$ である。中心極限定理から、 $\sqrt{N}(2e - 1) \xrightarrow{d} N(0, \text{Var}(2e))$ なので、 $x = N(2e - 1) \xrightarrow{d} N(0, 4N\text{Var}(e))$ である。ここで、

$$\text{Var}(e) = E[e^2] - E[e]^2 = E \left[E \left[\frac{1}{N^2} (N - \# \{d_i > \alpha N(2e - 1)\})^2 \mid e \right] \right] - E[e]^2$$

ここで、 $N - \# \{d_i > \alpha N(2e - 1)\} \sim Bin(N, F(\alpha N(2e - 1)))$ であり、 $b = N - \# \{d_i > \alpha N(2e - 1)\}$ とすると、

$$\begin{aligned} E[e^2] &= E \left[\frac{1}{N^2} (\text{Var}(b \mid e) + E[b \mid e]^2) \right] \\ &= E \left[\frac{F(\alpha N(2e - 1))(1 - F(\alpha N(2e - 1)))}{N} + F(\alpha N(2e - 1))^2 \right] \end{aligned}$$

以上より

$$4N\text{Var}(e) = 4E[F(\alpha N(2e-1))(1-F(\alpha N(2e-1)))] + \{4E[F(\alpha N(2e-1))^2] - 1\}N \\ - \frac{2F^{-1}\left(1 - \int_0^1 G(z)dz\right)}{\alpha} - \frac{F^{-1}\left(1 - \int_0^1 G(z)dz\right)}{\alpha N}$$

この極限をとると、

$$\sigma^{2*} = \lim_{N \rightarrow \infty} 4N\text{Var}(e) = \lim_{N \rightarrow \infty} 4E[F(\alpha N(2e-1))(1-F(\alpha N(2e-1)))] \\ + \lim_{N \rightarrow \infty} \{4E[F(\alpha N(2e-1))^2] - 1\}N \\ - \frac{2F^{-1}\left(1 - \int_0^1 G(z)dz\right)}{\alpha}$$

これが収束するのなら $x \xrightarrow{d} N(0, \sigma^{2*})$ であるので、 ϕ を標準正規分布の密度関数として、

$$g(z) = -\frac{1}{\alpha f(F^{-1}(z))} \phi\left(\frac{F^{-1}(z)}{\alpha \sigma^*}\right)$$

を得る。これを用いて実際のデータ e_t に対して最尤法を適用することができ、パラメータの α を推定できる。

3 シミュレーションデータでの推定

4 実証研究

入試日程の設定と満員電車

5 考えるべきこと

- σ^* は収束する？（しなそう） σ^* は推定することもできる（？）