

卒論テーマ候補：満員電車

池上 慧

2017 年 7 月 31 日

1 概要

ある駅から通勤する人々の行動を考える。 $a, b \in \mathbb{R}_+$ として、区間 $[a, b]$ を以下のように R の小区間に分割するような電車の発車時間が与えられているとする。

$$\Delta = \{[\delta_{t-1}, \delta_t) \mid a \leq \delta_i \leq b, i = 1, \dots, R\}$$

ここで電車の発車時間は δ_i で表現されている。 $\Delta_t \in \Delta$ は電車 $t-1$ が発車してから電車 t が発車するまでの時間であり、この間に改札を通過した通勤者は電車 t に乗るものとする。

改札の通過時刻から観察された、実際に区間 Δ_t に改札を通過した人の数を \tilde{N}_t と表記する。しかしこの観測される人数は各個人にとって本来望ましい行動の結果として得られているものではないことが予想される。なぜならば、通勤者はなるべく空いている電車に乗りたいという望みを持っているからだ。通勤電車に混雑という現象が伴っていない時には個人 i は自身にとって最適な時間である電車 t_i を選択し、区間 Δ_{t_i} に改札を通過すれば良い。しかし、現実には同じ駅から通勤電車に乗る人々の通勤時間帯はほぼ同じになるため、特定の時間帯のみ満員電車が発生してしまう。満員電車はすなわち乗客が多いためそれ自体が遅延の原因となる可能性も高く、また単純に人が密集した不快な空間であるために可能であれば避けたい存在である。

このような状況で人々は「大多数の人と共通である自分にとっての最適な通勤電車からなるべく離れないようにしながら、かつ乗客の少ない電車に乗りたい」という問題を解いていることが想定される。ここではこのような意思決定を人々が行っているか否かの検証と、そうであるならば本来はどの通勤時間を望んでいるのかの推定を考える。そして推定した本来の分布を元に「最適な」ダイヤグラムの作成方法を考える。

2 モデル

今、 n 個の電車が存在する状況を考える。個人が本来望ましい電車を選択した時に電車 t に乗る人数を N_t と表記する。ここで t を電車の発車時刻とみなす。個人はなるべく遅い時間を本来望ましい電車としてしていると仮定して、通勤電車の選択を望ましいものとは違うものにする場合は必ず元の電車よりも時間の早いものから選ぶとする。この時個人ごとに電車 t と t' とを選ぶ時の効用の差が以下のようにかけるとする。

$$u_i^t - u_i^{t'} = \alpha(N_i^t - N_i^{t'}) + \beta(t - t') + \epsilon_i^t - \epsilon_i^{t'}$$

ただし、 ϵ_i^t は第 1 種極値分布に従う確率変数で、個人と電車に対して独立であるとする。この時もともと各区間に属する人ごとに、その区間よりも前の区間についての選択確率がかける。区間 t が本来望ましいものである人がそれよりも前の区間 t' を選ぶ確率を $s_{t'}^t$ で表記することにする。

この時、実際に観測される各区間の人数は以下のようにかける。

$$\begin{cases} \hat{N}_1 = N_1 + \sum_{l=1}^{n-1} N_{1+l} s_{1+l}^{1+l} \\ \vdots \\ \hat{N}_m = N_m \left(1 - \sum_{l=1}^{m-1} s_l^m\right) + \sum_{l=1}^{n-m} N_{m+l} s_m^{m+l} \\ \vdots \\ \hat{N}_n = N_n \left(1 - \sum_{l=1}^{n-1} s_l^n\right) \end{cases} \quad (1)$$

この n 本の連立方程式を元の分布である $N_1 \dots N_n$ について解くことができ、以下のように記述できる。 $m = 1, \dots, n$ として、

$$s_m^m N_m = \left(\sum_{i=0}^n N_{m+i} \gamma_i^m \right) \quad (2)$$

$$\text{where } \gamma_i^m = - \sum_{j=0}^{i-1} \gamma_j^{m+i-j} \frac{s_m^{m+i-j}}{s_{m+i-j}^{m+i-j}} \quad (3)$$

これを行列で表記すると以下のようである。

$$\begin{pmatrix} s_n^n N_n \\ s_{n-1}^{n-1} N_{n-1} \\ s_{n-2}^{n-2} N_{n-2} \\ \vdots \\ s_2^2 N_2 \\ s_1^1 N_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ \gamma_1^{n-1} & 1 & 0 & \dots & 0 & 0 \\ \gamma_2^{n-2} & \gamma_1^{n-2} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n-2}^2 & \gamma_{n-3}^2 & \dots & 1 & 0 & 0 \\ \gamma_{n-1}^1 & \gamma_{n-2}^1 & \dots & \gamma_1^1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \hat{N}_n \\ N_{n-1} \\ N_{n-2} \\ \vdots \\ \hat{N}_2 \\ \hat{N}_1 \end{pmatrix} \quad (4)$$

この掛け合わせる行列を Γ_n と表記する。 Γ_n の k 列目の中で 0 ではないようその部分を $\Gamma_{n,k}$ と置く。すなわち以下のようである。

$$\Gamma_{n,k} = \begin{pmatrix} 1 \\ \gamma_1^{n-k} \\ \gamma_2^{n-k+1} \\ \vdots \\ \gamma_{n-k+1}^2 \\ \gamma_{n-k}^1 \end{pmatrix} \quad (5)$$

(3) 式を用いると $\Gamma_{n,k}$ の要素については以下のような関係式が成立していることがわかる。

$$\begin{pmatrix} \gamma_1^{n-k} \\ \gamma_2^{n-k+1} \\ \vdots \\ \gamma_{n-k+1}^2 \\ \gamma_{n-k}^1 \end{pmatrix} = - \begin{pmatrix} \frac{s_{n-k+1}^{n-k+1}}{s_{n-k+1}^{n-k+1}} & 0 & 0 & \dots & 0 \\ \frac{s_{n-k+1}^{n-k+1}}{s_{n-k+1}^{n-k+1}} & \frac{s_{n-k}^{n-k}}{s_{n-k}^{n-k}} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{s_{n-k+1}^{n-k+1}}{s_{n-k+1}^{n-k+1}} & \frac{s_{n-k}^{n-k}}{s_{n-k}^{n-k}} & \dots & \frac{s_3^3}{s_3^3} & 0 \\ \frac{s_{n-k+1}^{n-k+1}}{s_{n-k+1}^{n-k+1}} & \frac{s_{n-k}^{n-k}}{s_{n-k}^{n-k}} & \dots & \frac{s_3^3}{s_3^3} & \frac{s_2^2}{s_2^2} \end{pmatrix} \begin{pmatrix} 1 \\ \gamma_1^{n-k} \\ \vdots \\ \gamma_{n-k-2}^3 \\ \gamma_{n-k+1}^2 \end{pmatrix} \quad (6)$$

以上より、選択確率を用いて逐次的に計算することで Γ_n の各列を容易に求めることができ、そうして構成した Γ_n を掛け合わせることで観測された分布から元の分布を再現することができる。

効用に関するパラメータをまとめて θ と書く。ここでは $\theta = (\alpha, \beta)$ である。推定したい元の分布は (N_1, \dots, N_n) で記述できるが、 $\sum_t N_t = \sum_t \hat{N}_t$ より、 (N_1, \dots, N_{n-1}) が推定できれば十分である。よって推定したい要素は $2 + (n-1) = n+1$ 個ある。しかし (4) 式は n 本の連立方程式なので、観測された分布を代入しても $n+1$ 個の要素をただ一つに定めることはできない。すなわちこのままではこのモデルは識別できない。