集計 POS データを用いたランダム係数ロジットモデルによるベイズ推定

株式会社インテージ 武村敦　　筑波大学システム情報系 近藤文代

消費者の購買行動予測のための需要モデルの研究には詳細な非集計データの使用が好ましいが、その入手が困難な場合がある。そのため、集計データから潜在変数として個人選択データの推定研究が行われた。その代表例に係数の推定方法に一般化モーメント(GMM)法を使用した Berry et al.(1995)のランダム係数ロジットモデルがある。一方、Jiang et al.(2009)はロジット集計シェアモデルの係数推定にベイズ法を用い、GMM 法に比べ良い精度であることを示した。Musalem et al. (2009) (1) は独立サンプル(IS)と消費者パネル(CP)の 2 つの異なる需要システムについてベイズ法により推定方法を提案したが、IS のみ実証分析を行った。本研究ではそれに着目し、実際の市場での有用な CP での推定方法にて実証分析を行いその有効性を確かめる。

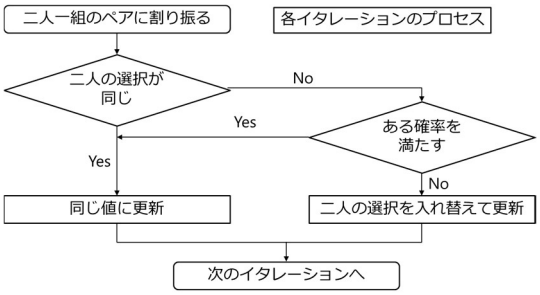
分析モデルは店舗 A の時点 t での消費者$i(i=1,\dots,I)$の商品購買$\boldsymbol{y}_{itj}$=1 に対応する選択肢jの選択確率$\Pr(y_{it}=j)$はp次元の説明変数ベクトルX_{ijt}及びその係数ベクトルβ_iから表される決定的成分$x_{ijt}\beta_i$を持つ多項ロジットモデルで表現される。個体間を規定する階層モデルはパネル固有変数w_iへ係数β_iを回帰させる形で定義され、$\boldsymbol{\Theta}$は階層係数ベクトルである。

次にデータ拡大法により非観測の消費者の個人選択を発生させるが、実際の観測集計市場シェア S_{jt}と一致する必要があり、制約条件$\sum_{i=1}^{N} z_{ijt} = NS_{jt}$が課される。その条件を満たす$z_{ijt}$は個人選択の組み合わせにより多く値が存在する。各組み合わせは各係数尤度および事後密度を持ち、各係数θ_iの事後分布だけでなく、個人の選択の組み合わせz_{ijt}に対応する分布も推定する必要がある。全てのモデル係数の事後密度は次式に比例して表される。

$$f(Z,\theta,\bar{\theta},D|S,X)$$

$$\propto\left(\prod_{i=1}^{N}\phi(\theta_i;\bar{\theta},D)\prod_{j=1}^{J}\prod_{t=1}^{T}I_{\{\sum_{i=1}^{N}z_{ijt}=NS_{jt}\}}P_{ijt}^{z_{ijt}}\right)$$

ここでZ,θ,S,Xは$z_{ijt},\theta_i,X_{ijt},S_{jt}$の行列である。消費者$i$の係数ベクトル$\theta_i$は MH 法を、$\theta_i$の分布の平均$\bar{\theta}$および分散共分散行列$D$はデータ拡大により推定された個人選択$z_{ijt}$を条件にギブスサンプリング法を用いて推定される。しかし、制約条件を満たすz_{ijt}の全ての可能な値の完全条件付き事後確率の計算は膨大となるため、消費者を二人一組に分けて個々の選択z_{ijt}を得てその解消を可能とした。(Step1) 各イタレーションk、各期 t に N 人のランダムに置き換えなしに選択された消費者から 2 人 1 組にわけ$\frac{N}{2}$個のペア pを作る。(Step2) 1 組目のペアから順に完全条件付き事後分布からペア $z_{i_{1pt}}^{k}, z_{i_{2pt}}^{k}$を同時抽出し、下図のステップに従ってある確率の下で個人選択の推定値を更新する。



分析には米国の大手スキャナデータ会社 IRI のデータを用い、ある店舗 A でのティッシュペーパー市場の主要 3 ブランドを対象とした。分析の結果、店舗 A のティッシュペーパー市場で、KL のブランド切片の推定値が他二つに比べ大きくなっており、観測データのブランドシェアの結果と一致した。

(1) Musalem, A., Bradlow, E. T., & Raju, J. S. (2009). Bayesian estimation of random‐coefficients choice models using aggregate data. Journal of Applied Econometrics, 24(3), 490-516.