離散選択モデルの最近の話題

名古屋大学　山本俊行

講義内容

- GEVモデル (generalized extreme value model)
- MMNLモデル (mixed multinomial logit model)
- 時間価値分布
- 離散連続モデル

GEVモデル

- 基礎
- モデルの拡張
- 新たな性質の発見

GEVモデル：基礎

- 多項ロジットモデルのIIA特性を緩和し，選択肢間の柔軟な誤差相関を可能としたモデル
  - 後述のMMNLモデルも同様の性質を有する
- 選択確率がclosed formで書けるため，数値積分の必要がない
  - 推定誤差の原因となる数値積分が必要なMMNLモデルより常に優先すべき
- あらゆる誤差相関に対応したGEVモデルが存在する訳ではない
  - 表現したい誤差相関を持つGEVモデルを見つけるか，見つからない時は自ら作る必要がある

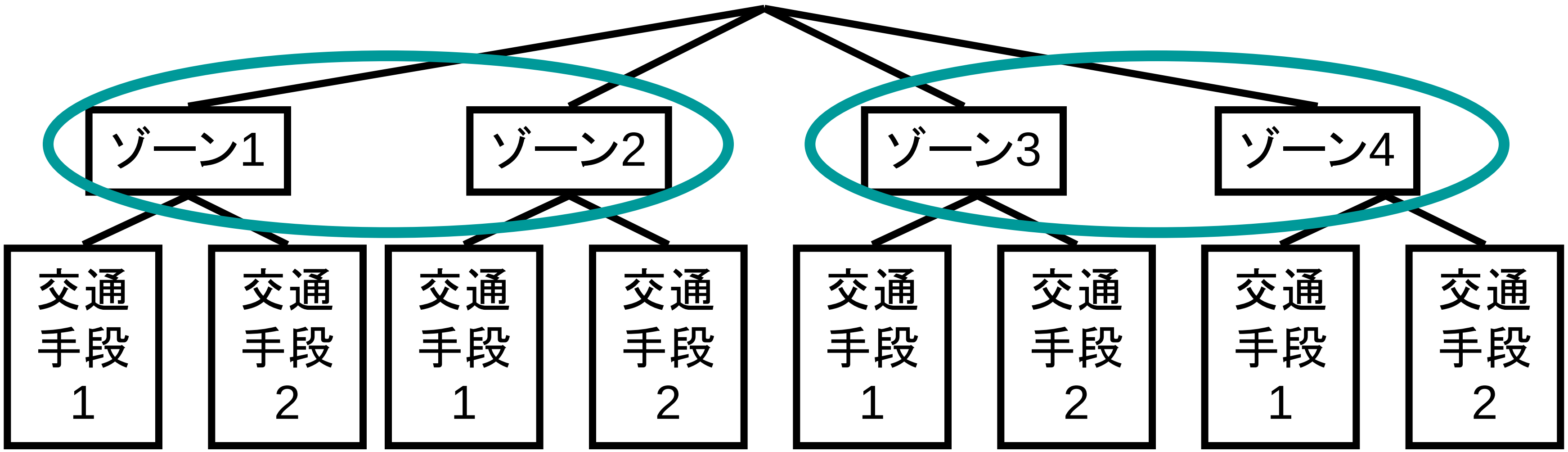
GEVモデル：モデルの拡張

- CNLモデルは等分散の誤差項間のあらゆる相関構造を表現できることが示された（Papola, 2004）
- GNLモデルがサンプル間の異分散と相関構造の異質性を表現できるように拡張された（Koppelman & Sethi, 2005）
- 選択肢間の相関構造をネットワーク構造で表現（Bierlaire, 2002）し，再帰型のネスティィッドロジットモデル（Daly, 2001）でモデル化することで，複雑な証明なしに新しいGEVモデルを開発する方法が提案された（Daly & Bierlaire, 2006）

GEVモデル：新たな性質の発見

ネスティッドロジットモデルの性質：

- 上位段階に目的地を持つ場合に選択肢（目的地）を統合しても整合的な効用関数が提案された（Ivanova, 2005）
  - もともとゾーンの設定は恣意的



GEVモデル：新たな性質の発見

ネスティッドロジットモデルの性質：

- 選択肢別抽出データを用いる場合，EMSL推定量が定数項を除き不偏性を持つことが示された（Garrow & Koppelman, 2005）
  - WESML推定量も不偏性を持つが有効性を持たない

- いずれも多項ロジットモデルで明らかであった性質をネスティッドロジットモデルに拡張している

MMNLモデル

- 基礎
- 数値積分法
- 効率的な数値積分
- 効率的なアルゴリズム
- MNPとの比較
- 効率的な推定方法

MMNLモデル：基礎

- 誤差項を多項ロジットモデルに追加したモデル
  - 選択肢間のあらゆる誤差分散共分散が表現可能
  - パラメータの非観測異質性も表現可能

$$P_{ni} = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \right) f(\beta)\, d\beta.$$

- 選択確率がclosed formで書けないため，数値積分が必要となる．シミュレーション法を用いる

$$\check{P}_{ni} = \frac{1}{R} \sum_{r=1}^{R} L_{ni}(\beta^r)$$

MMNLモデル：数値積分法

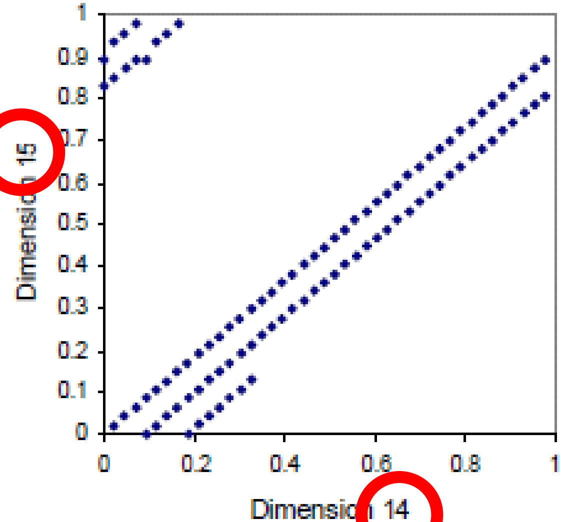
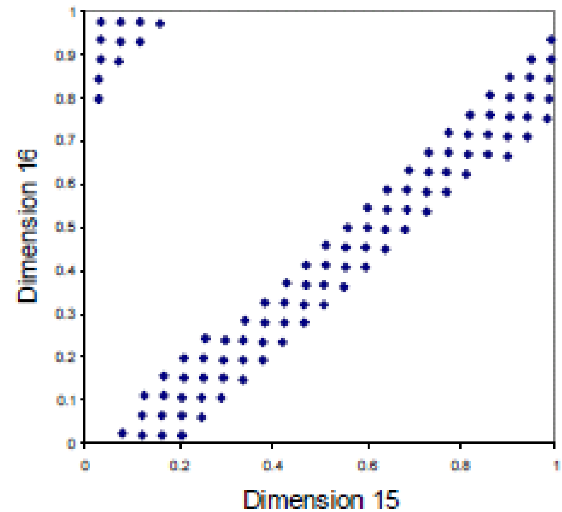
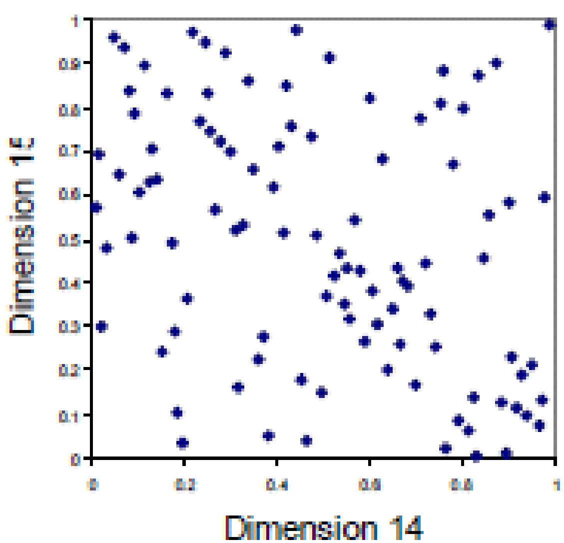
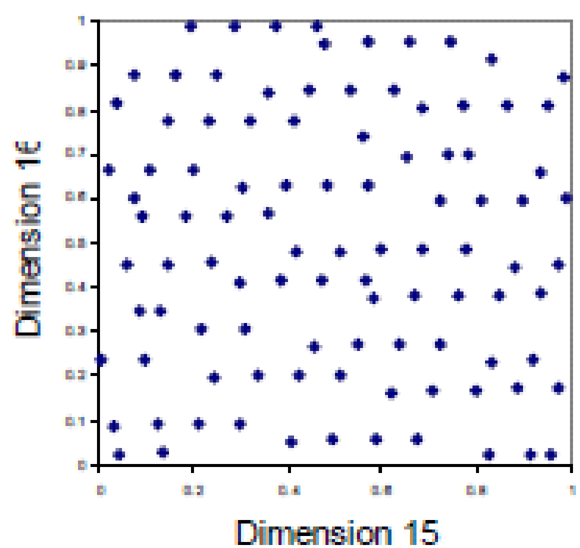
等間隔にシミュレーション点を取ると次元間の相関が生じ，推定結果にバイアスが生じる

- 擬似乱数法
  - 個々の点を乱数で決定する．ただし，計算機では真の乱数は発生できないため擬似乱数を用いる
- 準乱数法
  - 関数に従って個々の点を決定する．擬似乱数法より均質な分布を得ることが出来る
- ハイブリッド法
  - 準乱数法の点の配列を乱数によって乱数化する．次元間の配列の相関を防ぐ

MMNLモデル：効率的な数値積分

効率的なQuasi乱数列の提案

- (t, m, s)-netsはHalton数列より効率的（Sándor & Train, 2004）
- Halton数列とFaure数列（ (t, m, s)-netsの特殊形）および，それらのスクランブル版（ハイブリッド法）を比較した結果，Faure数列のハイブリッド版が最も効率的（Sivakumar, et al., 2005）
- MLHS (modified Latin hypercube sampling)はHalton数列，Halton数列のスクランブル版，シャッフル版よりも効率的（Hess, et al., 2006）
  - MLHSとFaure数列の比較はまだされていない

	Halton数列	Faure数列
標準型		
スクランブル		

Sivakumar et al. (2005)

MMNLモデル：効率的なアルゴリズム

BTRDA (basic trust-region with dynamic accuracy)アルゴリズム

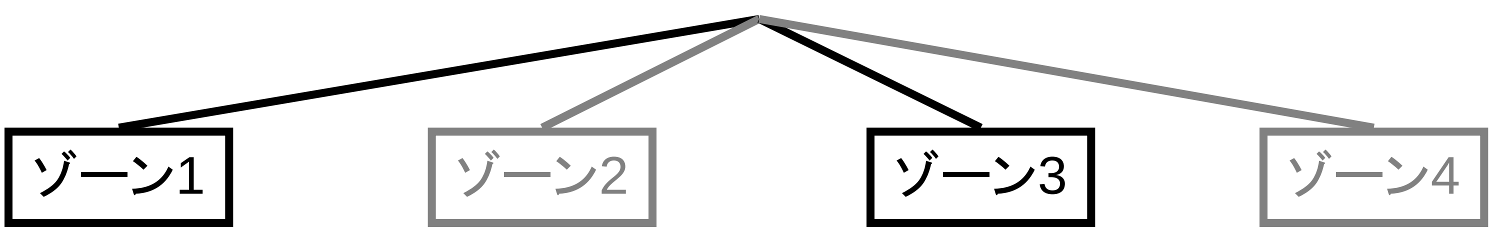
- 繰り返し計算の初期段階の抽出点数を削減し，計算速度の向上（Bastin, et al., 2006）
- BTRDAとMLHSの組み合わせはBFGSと純粋擬似乱数法の組み合わせよりも効率的（Bastin, et al., 2005）

MMNLモデル：MNPとの比較

- パネル分析（パラメータの非観測異質性と誤差項の系列相関を考慮）において，選択肢が25より小さい場合，MMNLモデルを擬似乱数列と組み合わせた場合よりもMNPモデルをGHKシミュレータと組み合わせた場合の方が効率的（Srinivasan & Mahmassani, 2005）
- 人工データを用いた分析の結果，MMNLモデルに擬似乱数法やHalton数列を組み合わせた場合，MNPモデルにGHKシミュレータを組み合わせた場合のいずれも，設定した誤差構造を再現するには8000程度のサンプル数が必要である（Minizaga & Alzarez-Dazian, 2005）

MMNLモデル：効率的な推定方法

- 推定時の選択肢サンプリング法
  - MNLモデルでは推定値の不偏性が証明されているが，MMNLモデルでは保持されない
  - MMNLモデルでは，4分の1から半分程度のサンプリングが安全である(Nerella & Bhat, 2004)
    - MNLモデルの場合は8分の1から4分の1



時間価値分布

- 基礎
- 分布形
- SPデータの信頼性

時間価値分布：基礎

- プロジェクト評価等に用いられる重要な指標
- 離散選択モデルの重要な出力の一つ
- 線形効用関数の場合，所要時間のパラメータを費用のパラメータで除した値が時間価値となる

$$V_i = \alpha_i - \beta_c C_i - \beta_t T_i \qquad \frac{\partial V_i / \partial T_i}{\partial V_i / \partial C_i} = \frac{\beta_t}{\beta_c}$$

- 時間のパラメータをランダム係数にすると時間価値も分布を持つ

時間価値分布:分布形

- 通常のランダム係数モデルでは分布形として正規分布を仮定することが多いが，時間価値が負の部分が生じてしまうので問題がある
- 問題の生じない分布形が提案されている
  - 切断正規分布，対数正規分布，一様分布，三角分布，Johnson's S_B 分布

$$c = a + (b-a) \cdot \frac{\exp(\xi)}{1+\exp(\xi)}, \quad \xi \sim N(\mu, \sigma)$$

- 時間価値分布のノンパラメトリック，セミパラメトリックな推定も試みられている（Fosgerau, 2006）
- 分散の非観測異質性を考慮するとモデルの精度が向上する（Greene, et al., 2006）

時間価値分布：SPデータの信頼性

- 既存研究のレビューの結果，SPデータを用いた場合に時間価値が過小推計される（Brownstone & Small, 2005）
- SP調査のデザインが被験者の意思決定過程に影響を及ぼすため，SPデザインの次元を考慮しないと時間価値は過小推計となる（Hensher, 2006）

離散連続モデル

- 基礎
- モデルの拡張

離散連続モデル：基礎

- 離散的な選択だけではなく，連続的な量の選択も効用理論に基づいたモデルが望まれる
- 離散的な選択と連続量の選択を同時に取り扱うモデルを離散連続モデルと呼ぶ
  - 自動車の車種選択と走行距離，暖房器具の選択と使用量，電話料金プランと利用量，等

離散連続モデル：モデルの拡張

- 同時に複数の財を選択し，それぞれの連続量も選択する場合のモデルへの拡張
  - 活動内容選択と活動時間，複数台保有世帯の自動車の車種選択と走行距離

- Kim, et al. (2002)は古典的な離散連続モデルを複数選択肢の同時選択に拡張した．誤差項に正規分布を仮定し，GHKシミュレーターによって多次元正規分布を積分し，メトロポリス法によるベイズ推定を行った

- Bhat (2005)はKim et al. (2002)と同様の構造の誤差項にガンベル分布を仮定し，スクランブルHalton数列によるシミュレーション法を用いた推定方法を提案した

終わりに

これからも論文は書けそう．．．

だが，

社会的要請はあるのか？？？