

解析的積分近似に基づく プロビット型鉄道経路選択モデルの計算特性

亀谷 淳平¹・福田 大輔²

¹学生会員 東京工業大学大学院修士課程 理工学研究科土木工学専攻
(〒152-8552 東京都目黒区大岡山2-12-1)

E-mail:j.kamegai@plan.cv.titech.ac.jp

²正会員 東京工業大学大学院准教授 理工学研究科土木工学専攻 (〒152-8552 東京都目黒区大岡山2-12-1)

E-mail:fukuda@plan.cv.titech.ac.jp

高密度な鉄道ネットワークにおける旅客の経路選択行動の記述に用いられてきた多項プロビットモデルでは、モデルに含まれる多変量正規分布の積分計算を陽に解くことができないため、モンテカルロ法に基づく近似計算が主に適用されてきた。一方鉄道需要予測の実務においては、より細密な旅客需要配分が必要とされており、精度要件を満足した上での計算速度向上が求められている。本研究は、4種類の解析的積分近似計算手法をプロビットモデルに適用し、数値解析を通じて各手法の特徴を明らかにした。また、首都圏鉄道ネットワークの旅客配分計算への適用を行い、精度並びに計算効率について従来手法との比較を行った。以上を通じて、解析的積分近似手法の一定の有用性を明らかにした。

Key Word : *multinomial probit model, analytical approximation, rail-route choice analysis, transit assignment, multivariate normal distribution*

1. はじめに

首都圏鉄道ネットワークのような高密度なネットワークでは同一の発着地点に対して複数の経路が存在し、利用者はこれらの複数の経路から選択を行う。また、高密度ネットワークの場合には、著しい経路の重複が発生する。このような交通ネットワークの整備効果を評価する際には旅客需要の変化等を的確に予測できる鉄道需要予測が必要となる。鉄道需要予測において特に重要となるのが旅客配分計算であり、首都圏鉄道ネットワークのような高密度なネットワークにおける旅客の選択行動の記述には、経路間の類似性を表現できる多項プロビットモデルが実務でも用いられてきた¹⁾。

しかし、プロビットモデルに含まれる多変量正規分布の積分計算は、解析的に解くことができない。そのため、現在までに近似計算に関する多くの研究が行われてきた。1960年代には解析的な積分近似に基づく手法が主流であった。しかし、当時の近似方法は一般に精度は高くなく、選択肢数も最大5個までしか扱えなかった²⁾。1990年代にはモンテカルロシミュレーションによる近似計算が進展した。その代表的なものとしてGHKシミュレーター³⁾⁴⁾⁵⁾が挙げられる。そして近年では、解析的な積分近似手

法が再び注目を集めている。その理由の1つに、GHKシミュレーターによる近似計算は膨大な量の乱数を生成して計算を行うため、計算速度に限界があることが挙げられる。また別の理由として、解析的な積分近似手法の計算精度の大幅な向上⁷⁾がなされていることも挙げられる。

さて、近年の鉄道需要予測の実務においては、時代情勢に伴い予測の詳細化が求められている。従来の鉄道需要予測では年齢や性別、時間帯の違いによる鉄道の利用形態の違いを考慮せず、すべての旅客を同一のものとして分析を行っていた⁸⁾。しかし、近年の高齢化社会の進展や、女性の社会進出等により、生活様式の多様化が促進され、これらの違いに伴う旅客行動の変化が顕著となった。このような、性年齢層別の違い、時間帯の行動の違い等、さまざまな条件ごとに鉄道需要予測の細分化を行う必要性が高まっている。さらに、それらのシナリオを組み合わせ、膨大な量の配分計算を行い政策シミュレーションを行う必要があることから、一定の精度を確保した上でのプロビットモデル計算速度のさらなる高速化が求められている。

以上の問題意識のもと、本研究は近年展開が進んでいる4種類の解析的積分近似計算手法をプロビットモデルに適用し、各手法の計算速度、精度に関する特徴の把握

を行うことにより、それらの有用性を検討することを目的とする。まず、簡単なネットワークを想定し4種類の手法を適用した配分計算を行い、各手法の検証をする。次に、実際の首都圏鉄道ネットワークの旅客配分計算への適用を行い、精度並びに計算効率について従来手法との比較を通じて、その有用性を検討する。

2. 解析的積分近似手法の概要

まず、プロビットモデルの基本形を導出する。個人が各選択肢から得る効用 U は、観測可能な部分（確定項） V と観測不可能な部分（誤差項） e の線形和からなる式(1)によって表現される。

$$U = V + e \quad (1)$$

ここで各選択肢の誤差項に正規分布を仮定することにより、プロビットモデルが導出される。 K 個の選択肢の中から j 番目の選択肢を選ぶ確率 P_j は効用差から以下の式(2)によって表される。

$$P_j = P_r [e_j - e_k > V_k - V_j | \forall k \in K; k \neq j] \\ = P_r [Z_j < z_j | \forall k \in K; k \neq j] \quad (2)$$

$$Z_k = e_k - e_j, z_k = V_j - V_k$$

この式(2)の確率は以下の式(3)によって求められる。

$$P = \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_{K-1}} \frac{1}{\sqrt{(2\pi)^{K-1} |\Sigma_z|}} \exp\left[-\frac{1}{2} \mathbf{Z}^T \Sigma_z^{-1} \mathbf{Z}\right] dZ_1 \cdots dZ_{K-1} \quad (3)$$

ここで、 Σ_z は誤差項の差の分散共分散行列である。

本研究では、上記のプロビットモデルを用いた選択確率計算に対して、Variable reordering⁹⁾, Separation-of-variable methods^{11), 12)}, Tridiagonal decomposition methods^{13), 14), 15)}, Spherical-radial transformation methods^{16), 17), 18), 19)}という4種類の解析的積分近似手法を適用する。

(1) Variable Reordering

多変量正規分布の確率計算を行う際に、積分範囲と分散共分散行列の対応する行と列を並び替えることによって、計算結果を変えずに確率計算に用いる値を並び替えることができる。

プロビットモデルを用いた選択確率 P は条件付確率の積で表すことができ、以下の式で示される。

$$P = P(Z_1 < z_1) \times P(Z_2 < z_2 | Z_1 < z_1) \times \cdots \\ \times P(Z_{K-1} < z_{K-1} | Z_1 < z_1, \dots, Z_{K-2} < z_{K-2}) \quad (4)$$

上記の式(4)の1つ目の確率 $P(Z_1 < z_1)$ が最小となるように

$$i = \arg \min_{1 \leq i \leq K-1} \left\{ \Phi\left(z_i / \sqrt{\sigma_{i,i}}\right) \right\}$$

となる i を求める。ここで Φ は標準正規分布の累積密度関数を表し、 $\Sigma_z = (\sigma_{i,j})$ である。この i を用いて積分範囲、分散共分散行列 Σ_z の行と列の1番目と i 番目を入れ替えることにより、

$$Z_1 = \frac{1}{\Phi(z_1)} \int_{-\infty}^{z_1} s \phi(s) ds$$

と仮定する。

この操作を繰り返し行い、 j 番目の条件付確率 $P(Z_j < z_j | Z_1 < z_1, \dots, Z_{j-1} < z_{j-1})$ についても同様に分散共分散行列 Σ_z をコレスキー分解した $\mathbf{L} = (l_{i,j})$ を用いて、

$$i = \arg \min_{j \leq i \leq K-1} \left\{ \Phi\left(\frac{z_i - \sum_{m=1}^{j-1} l_{i,m} y_m}{\sqrt{\sigma_{i,i} - \sum_{m=1}^{j-1} l_{i,m}^2}}\right) \right\}$$

となるような i を求めて積分範囲、分散共分散行列 Σ_z の行と列の j 番目と i 番目を入れ替え、

$$Z_j = \frac{1}{\Phi(z_j)} \int_{-\infty}^{z_j} s \phi(s) ds$$

$$\tilde{z}_j = \frac{z_j - \sum_{m=1}^{j-1} l_{j,m} Z_m}{l_{j,j}}$$

と仮定する。

この操作を $K-2$ 回繰り返すことで求める確率は以下の式(5)で表現される。

$$P = \int_{-\infty}^{\tilde{z}_1} \phi(y_1) \int_{-\infty}^{\tilde{z}_2(Z_1)} \phi(y_2) \cdots \int_{-\infty}^{\tilde{z}_{K-1}(Z_1, \dots, Z_{K-2})} \phi(y_{K-1}) dy \quad (5)$$

式(5)は一次元の標準正規分布の積であり、以下の式(6)のように表される。

$$P = \Phi(\tilde{z}_1) \times \Phi(\tilde{z}_2(Z_1)) \times \cdots \times \Phi(\tilde{z}_{K-1}(Z_1, \dots, Z_{K-2})) \quad (6)$$

(2) Separation-of-Variable Methods

この手法は、 Z_i を Z_1, \dots, Z_{i-1} の関数として評価する。式変形後、最も内側の積分は Z_1, \dots, Z_{K-1} を積分変数とし、最も外側の積分は Z_1 のみを積分変数とする。さらに、式変形後は積分区間を $[0,1]^{K-1}$ に標準化する。

分散共分散行列 Σ_z をコレスキー分解($\Sigma_z = \mathbf{L}\mathbf{L}^T$)した後の多変量正規分布の確率は式(7)となる。

$$P = \int_{-\infty}^{\tilde{z}_1} \phi(y_1) \int_{-\infty}^{\tilde{z}_2(y_1)} \phi(y_2) \cdots \int_{-\infty}^{\tilde{z}_{K-1}(y_1, \dots, y_{K-1})} \phi(y_{K-1}) dy \quad (7)$$

$$z_i = \frac{z_i - \sum_{j=1}^{i-1} l_{i,j} y_j}{l_{i,i}}$$

ここで、 $y_i = \Phi^{-1}(x_i)$ と置くと式(7)は以下の式(8)のようになる。

$$P = \int_{-\infty}^{\Phi(\tilde{z}_1)} \int_{-\infty}^{\Phi(\tilde{z}_2(\Phi^{-1}(x_1)))} \dots \int_{-\infty}^{\Phi(\tilde{z}_{K-1}(\Phi^{-1}(x_1), \dots, \Phi^{-1}(x_{K-2})))} \mathbf{dx} \quad (8)$$

最後に

$$e_i(w_1, \dots, w_{i-1}) = \Phi(\tilde{z}_i(\Phi^{-1}(z_1(w_1)), \dots, \Phi^{-1}(z_{i-1}(w_{i-1}))))$$

を用いて $x_i = e_i w_i$, $\mathbf{dx}_i = e_i \mathbf{dw}_i$ と置くと、求める選択

確率は以下の式(9)で示される。

$$P = e_1 \int_0^1 e_2(w_1) \dots \int_0^1 e_{K-1}(w_1, \dots, w_{K-2}) \int_0^1 \mathbf{dw} \quad (9)$$

(3) Tridiagonal Decomposition Methods

分散共分散行列が三重対角行列のような帯行列となる時、多変量正規分布は計算コストを小さくすることができる。これに着目したのが本手法である。

分散共分散行列 Σ_z をコレスキー分解し、正規化を行うと多変量正規分布の累積密度関数は

$$\Phi(\mathbf{Z} < \mathbf{z}; \Sigma_z) = \Phi(\mathbf{L}\mathbf{y} < \mathbf{z}; \mathbf{I})$$

で示される。コレスキー行列 \mathbf{L} をさらに分解することによって \mathbf{L} を下二重行列に変形させることが、この手法の基本的な考えである。まず、 $\mathbf{L}\mathbf{Q}$ が以下のような直交行列 \mathbf{Q} を設定する。

$$\mathbf{L}\mathbf{Q} = \begin{bmatrix} * & 0 & \dots & 0 & 0 \\ ? & * & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ ? & \dots & ? & * & 0 \\ ? & ? & \dots & ? & * \end{bmatrix} \mathbf{Q} = \begin{bmatrix} ? & \dots & \dots & ? & 0 \\ ? & \dots & \dots & ? & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ ? & \dots & \dots & ? & 0 \\ 0 & \dots & 0 & ? & * \end{bmatrix}$$

上記の行列の*は0ではない値を表しており、?は0ではない場合と0である場合のどちらもとり得る値を示している。

$\mathbf{x} = \mathbf{Q}'\mathbf{y}$ と置くと多変量正規分布の累積密度関数は

$$\Phi(\mathbf{Z} < \mathbf{z}; \Sigma_z) = \Phi(\mathbf{L}\mathbf{Q}\mathbf{x} < \mathbf{z}; \mathbf{I})$$

で示される。ここで、凸包計算アルゴリズムの一つである Fourier-Motzkin elimination を用いると $\mathbf{L}\mathbf{Q}\mathbf{x} < \mathbf{z}$ は $K-2$ 通

りの $\mathbf{M}_j^{(1)} \mathbf{x} < \mathbf{z}_j$ に分解できる。 $\mathbf{M}_j^{(1)}$ は以下のような行列である。

$$\mathbf{M}_j^{(1)} = \begin{bmatrix} ? & ? & \dots & ? & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ ? & ? & \dots & ? & 0 & 0 \\ ? & ? & \dots & ? & * & 0 \\ 0 & 0 & \dots & 0 & ? & * \end{bmatrix} \text{for } (j=1, \dots, K-2)$$

このときの多変量正規分布の確率は以下の式(10)で示される。

$$P = \sum_{j=1}^{K-2} \int_{-\infty < \mathbf{M}_j^{(1)} \mathbf{x} < \mathbf{z}_j} \frac{e^{-\frac{\mathbf{x}\mathbf{x}}{2}}}{\sqrt{(2\pi)^{K-1}}} \mathbf{dx} \quad (10)$$

この操作を $K-2$ 回繰り返し、 $(K-2)!$ 通りの $\mathbf{M}_j^{(K-2)} \mathbf{x} < \mathbf{z}_j$ に分解する。 $\mathbf{M}_j^{(K-2)}$ は対角成分がすべて1になるように標準化することで以下のような行列となる。

$$\mathbf{M}_j^{(K-2)} = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ c_1 & 1 & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & c_{K-2} & 1 \end{bmatrix} \text{for } (j=1, \dots, (K-2)!)$$

このときの多変量正規分布の確率は式(11)で示される。

$$P = \sum_{j=1}^{(K-2)!} \int_{-\infty < \mathbf{M}_j^{(K-2)} \mathbf{x} < \mathbf{z}_j} \frac{e^{-\frac{\mathbf{x}\mathbf{x}}{2}}}{\sqrt{(2\pi)^{K-1}}} \mathbf{dx} \quad (11)$$

$$= \sum_{j=1}^{(K-2)!} \Phi(\mathbf{Z} < \mathbf{z}_j; \mathbf{T}_j)$$

式(11)の \mathbf{T}_j は $\mathbf{T}_j = \mathbf{M}_j^{(K-2)} (\mathbf{M}_j^{(K-2)})'$ となる三重対角行列であり、再帰的に求めることができる。

$$\Phi(\mathbf{Z} < \mathbf{z}_j; \mathbf{T}_j) = \int_{-\infty}^{z_{j1}} \phi(y_1) \int_{-\infty}^{z_{j2}-c_{11}y_1} \phi(y_2) \dots \int_{-\infty}^{z_{j,K-1}-c_{K-2}y_{K-2}} \phi(y_{K-1}) \mathbf{dy} \quad (12)$$

となるので、

$$g_{K-1}(y) = \Phi(z_{j,K-1} - c_{K-2}y)$$

$$g_i(y) = \int_{-\infty}^{z_{ji}-c_{i-1}y} \phi(t) g_{i+1}(t) \mathbf{dt}$$

仮定し、 $i = K-2, K-3, \dots, 1$ を代入すると式(12)は以下の式(13)となる。

$$\Phi(\mathbf{Z} < \mathbf{z}_j; \mathbf{T}_j) = g_1 = \int_{-\infty}^{z_{j1}} \phi(t) g_2(t) \mathbf{dt} \quad (13)$$

(4) Spherical-Radial Transformation Methods

この手法は極座標表示を行う近似計算である。まず、コレスキー分解後の多変量正規分布の確率は、式(14)となる。

$$P = \int_{-\infty}^{\tilde{z}_1} \phi(y_1) \int_{-\infty}^{\tilde{z}_2(y_1)} \phi(y_2) \dots \int_{-\infty}^{\tilde{z}_{K-1}(y_1, \dots, y_{K-1})} \phi(y_{K-1}) \mathbf{dy} \quad (14)$$

ここで、ベクトル \mathbf{y} を動径 r と方向ベクトル \mathbf{x} を用いて極座標変換する。動径 r と方向ベクトル \mathbf{x} を用いると式(14)は以下のように変形される。

$$P = \int_{\|\mathbf{x}\|=1} \frac{2^{1-K/2}}{\Gamma(K/2)} \int_{r_1(\mathbf{x})}^{r_u(\mathbf{x})} r^{K-2} e^{-\frac{r^2}{2}} dr dU(\mathbf{x}) \quad (15)$$

式(15)の $U(\mathbf{x})$ は \mathbf{x} 上の同時累積分布関数を表している。

$$R(\mathbf{x}) = \{r : r \geq 0, r\mathbf{L}\mathbf{x} < \mathbf{z}\}$$

とすると、式の内側の積分の積分区間は、

$$r_l(\mathbf{x}) = \min\{r : r \in R(\mathbf{x})\}$$

$$r_u(\mathbf{x}) = \max\{r : r \in R(\mathbf{x})\}$$

で定義される。さらに、 $\mathbf{V}=\mathbf{L}\mathbf{x}$ と置くと積分区間は以下のように求めることができる。

$$r_l(\mathbf{x}) = \max\left\{0, \max_{v_i < 0} \left\{\frac{z_i}{v_i}\right\}\right\}$$

$$r_u(\mathbf{x}) = \max\left\{0, \min_{v_i > 0} \left\{\frac{z_i}{v_i}\right\}\right\}$$

(5) 4つの解析的積分近似手法の特徴整理

まず、計算手法について比較を行う。計算手法は3つに大別することができる。1つ目のグループはSpherical-Radial Transformation Methodsである。この手法の最大の特徴は座標系を直交座標から極座標に変換させることである。

2つ目はSeparation-of-Variables MethodsとTridiagonal Decomposition Methodsである。Separation-of-Variables Methodsは Z_i を Z_1, \dots, Z_{K-1} の関数で評価し、Tridiagonal Decomposition Methodsはこの関数で評価して積分近似を行っている。

3つ目は、Variable Reorderingであり、この手法は最終的に次元の標準正規分布の積で表現することができる。これはGHKシミュレーターを用いた際の近似計算に類似している。

これらの4手法の中で、計算速度が一番速い手法はVariable Reorderingである。この手法は最終的に次元の標準正規分布の積で表現できるため計算コストが小さくなる。逆に、最も計算速度が遅い手法は、Tridiagonal Decomposition Methodsである。この手法は多変量正規分布の累積分布関数を $(K-2)!$ 個に分解し、それぞれの多変量正規分布の確率を求めなければならないため計算コストが大きくなる。一方、計算コストが大きくなるのに伴い計算精度も高くなるので、最も計算速度が遅いTridiagonal Decomposition Methodsは計算精度が最も高くなり、逆にVariable Reorderingは計算精度が低くなる。

3. 仮想ネットワークを用いた精度検証

(1) 選択肢4つのネットワーク

a) ネットワークの設定

簡単なネットワークを想定し、4つの手法の精度検証を行い、その特徴を考察する。今、4つの経路が存在するネットワークが存在し、4つの経路がすべて独立なケース (図-1)、経路3と経路4が相関を持つケース (図-2)

を想定する。このときの分散共分散行列は以下のように表現する。

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \Sigma' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0.9 & 1 \end{bmatrix}$$

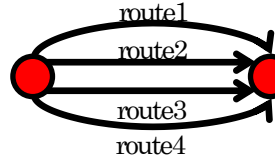


図-1 独立なケース

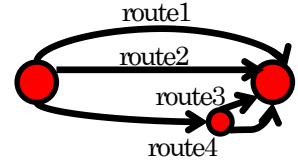


図-2 相関のあるケース

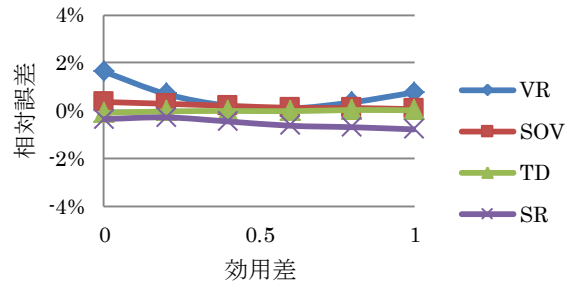


図-3 4経路が独立な場合の選択確率計算結果

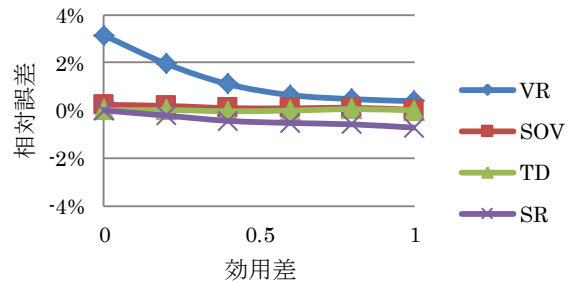


図-4 3と4に相関のある場合の選択確率計算結果

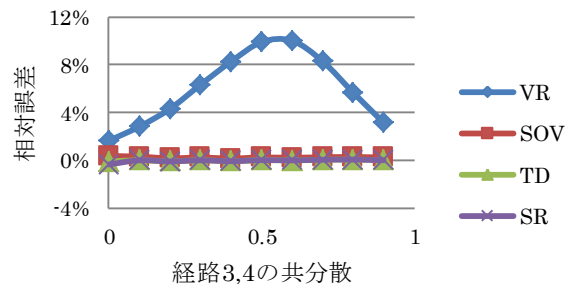


図-5 相関が変化した場合の選択確率計算結果

b) 選択確率計算結果と考察

本研究では、プロビットモデルによる選択確率計算のコードにR言語を使用している。解析的積分近似に基づく計算手法のいずれも最終的に次元関数の積分を行わなければならない。本研究ではその次元の積分を、合成シンプソン公式による数値積分を用いて解く。その際に積分区間の分割数を10とする。また、GHKシミュレーターによる選択確率の計算は、乱数の生成数を非常に大きくすることで精度は上がり、真値へと近づいていく。ここでは、乱数生成を100,000回行った際のGHKシミュレーターによる選択確率計算結果を真値と見立てて各解析的近似手法との比較を行い考察する。

図-3に4つの経路がすべて独立な場合の結果を、図-4に経路3と経路4が相関を持つ場合の結果を、図-5に経路3と経路4の相関が変化した場合の結果を示す。

まず、効用差の違いについて考察する。図-3、図-4を見ると4つの手法すべてでGHKシミュレーターによる計算結果との相対誤差が十分に小さい値となった。したがって、選択枝の数が少ない場合には、効用差の違いによる影響はほとんど見られず、4つの手法すべてで十分な精度の近似ができることを示唆している。1番誤差が大きくなったのはVariable Reorderingである。ただし、その最大値は相関のあるケースの効用差が0のときの3.14%と精度は高い。2番目に誤差が大きくなったのはSpherical-Radial transformation methodsであるが、その値はすべての場合で1%未満と非常に小さい。Separation-of-Variable Methods, Tridiagonal Decomposition Methodsはすべての場合で相対誤差が0.5%未満となり、非常に高い精度を保っていた。以上より、選択枝が少ない場合には4つの手法すべてで十分な精度の近似をすることができることが明らかとなった。また、効用差の違いによって誤差の大きさも変動するが、その変動はわずかなものであり、近時計算の誤差に与える影響が小さいことも示唆された。

次に、相関の違いについて考察する。図-5を見るとVariable Reorderingのみが真値よりも大きな値をとっている。特に、共分散が0.5, 0.6のときには相対誤差がおおよそ10%と大きい。つまり、Variable Reorderingは経路間に中程度の相関がある際に選択確率を真値よりも大きく評価する傾向がある。一方、残りの3つの手法は相関の変化による影響がほとんど見られず、相関がある場合でも十分な精度で近似できることが明らかとなった。

(2) 選択枝数の増加による影響

a) ネットワークの設定

次にネットワークの経路数を4から10まで変化させて、選択枝数の増加による影響を検証を行い、その特徴を考察する。今、すべての経路が独立なケース(図-6)、すべての経路が互いに相関を持っているケース(図-7)を

想定する。相関を持つ場合の分散共分散行列は以下のよう表現する。

$$\Sigma = \begin{bmatrix} 1 & \frac{K-2}{K} & \frac{K-3}{K} & \dots & \frac{1}{K} & 0 \\ \frac{K-2}{K} & 1 & \frac{K-2}{K} & \dots & \frac{2}{K} & \frac{1}{K} \\ \frac{K-3}{K} & \frac{K-2}{K} & \ddots & \ddots & \ddots & \frac{2}{K} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \frac{1}{K} & \frac{2}{K} & \ddots & \ddots & 1 & \frac{K-2}{K} \\ 0 & \frac{1}{K} & \frac{2}{K} & \dots & \frac{K-2}{K} & 1 \end{bmatrix}$$

b) 選択確率計算結果と考察

選択枝が5つ以上の場合のプロビットモデルによる選択確率計算を行い、選択枝が4つの場合と同様に乱数生成数が100,000回の際のGHKシミュレーターによる選択確率計算結果を真値と見立てて各解析的近似手法との比較を行い考察する。ただし、ここではVariable ReorderingとSeparation-of-Variable Methodsの2つの手法のみを扱う。

図-8にすべての経路が独立な場合の結果を、図-9にす

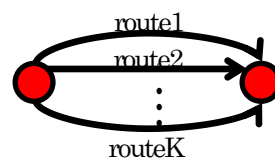


図-6 独立なケース(複数)

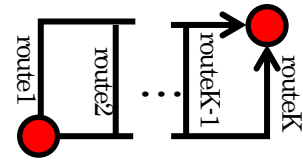


図-7 相関のあるケース(複数)

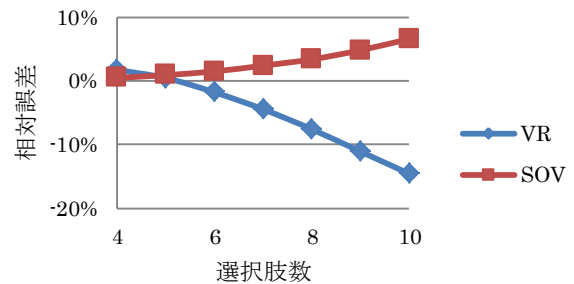


図-8 独立なケース(複数)の選択確率計算結果

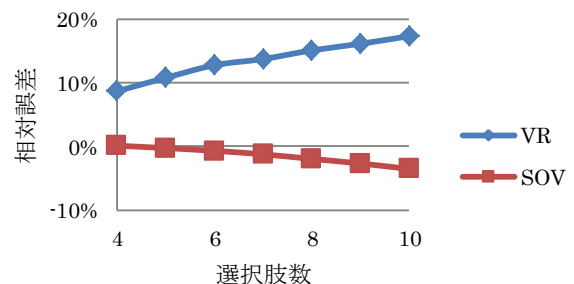


図-9 相関のあるケース(複数)の選択確率計算結果

すべての経路が互いに相関を持つ場合の結果を示す。

選択枝数が少ない場合には、Variable ReorderingとSeparation-of-Variable Methodsの2つの手法ともに誤差が小さいが、選択枝数の増加に伴い誤差が大きくなった。その増加量はVariable Reorderingを用いた際に特に大きくなり、すべての経路が独立で選択枝が10個のときには相対誤差が14.5%、経路間に相関があるときには相対誤差が17.2%となった。一方、Separation-of-Variable Methodsを用いた際にはすべての経路が独立な場合は、選択枝が10個のときに相対誤差が最大となり、その値は6.56%、経路間に相関があるときも選択枝が10個のときに相対誤差が最大となり、その値は3.57%であり、選択枝の増加に伴い誤差は大きくなるが、その増加量はVariable Reorderingよりも小さい。

以上より、選択枝が少ない場合は、Variable ReorderingとSeparation-of-Variable Methodsのいずれの手法を用いても高い精度で近似できたが、選択枝の増加に伴い誤差も大きくなり精度が低くなることが明らかとなった。

(3) 計算精度と速度のトレードオフ関係

4種類の解析的な積分近似手法は、いずれも最終的に一次元関数の積分の積に帰着する。本研究では、この積分を合成シンプソン公式を用いた数値積分によって求めているが、このときの数値積分の精度を上げることでプロビットモデルによる選択確率の計算精度も上がる。一方、数値積分の精度を上げるためには積分区間をより細かく分割する必要があるが、それに伴い計算に要する時間も大きくなる。つまり、選択確率を求める際には、選択確率計算精度と計算速度の間にトレードオフの関係が存在する。そこで、それぞれの手法の数値積分の精度と計算速度のトレードオフ関係を検証する。

プロビットモデルによる選択確率計算精度と計算速度のトレードオフ関係を検証するために合成シンプソン公式の積分区間の分割数 n を $n=10,20,50,100$ と増加させて選択確率を計算する。選択枝が4つでそのすべての選択枝の効用が等しい互いに独立なネットワークを想定し経路1を選択する場合の選択確率を求める。このとき、4つの

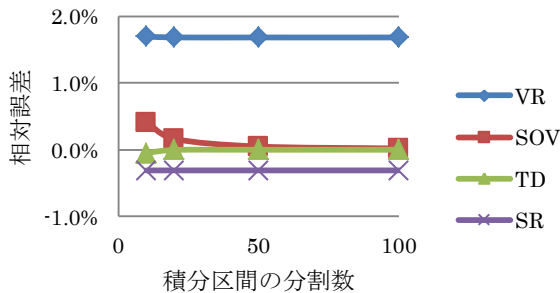


図-10 積分区間の分割数が増加した場合の選択確率計算結果

経路の選択確率はすべて等しいため、経路1の選択確率の真値は0.25となる。

図-10に積分区間の分割数を増加させた際の結果を示す。また、それぞれの手法の数値積分の精度と計算速度の関係を示し、上記の結果とあわせて考察を行う。

まず、Variable Reorderingは以下の式(16)によって選択確率を求めている。

$$P = \Phi(\tilde{z}_1) \times \Phi(\tilde{z}_2(Z_1)) \times \dots \times \Phi(\tilde{z}_{K-1}(Z_1, \dots, Z_{K-2})) \quad (16)$$

$$Z_j = \frac{1}{\Phi(\tilde{z}_j)} \int_{-\infty}^{\tilde{z}_j} s \phi(s) ds$$

ここで、合成シンプソン公式の積分区間の分割数を2倍にすると式(16)の積分にかかるコストも2倍に、分割数を3倍にすると式(16)の積分コストも3倍にといったように、等倍で積分コストが大きくなる。

一方、図-10に注目すると、相対誤差はほぼ一定の値をとっていることがわかる。これは、数値積分の精度が十分であれば計算結果にあまり影響を与えないことを示唆している。つまり、Variable Reorderingは数値積分の精度をあげても真値に近づくわけではなく、ある値に収束するというを示している。以上より、数値積分の精度を上げると、計算にかかる時間が増加するが、プロビット確率はある一定の値で収束しそれ以上真値に近づくことはないことが示唆される。

Separation-of-Variable Methodsは以下の式(17)によって選択確率を求めている。

$$P = e_1 \int_0^1 e_2(w_1) \dots \int_0^1 e_{K-1}(w_1, \dots, w_{K-2}) \int_0^1 dw \quad (17)$$

Separation-of-Variables Methodsは積分の中に積分が存在する構造であり、合成シンプソン公式の積分区間の分割数を2倍にすると1番内側の積分コストは2倍に、内側から2番目の積分コストは4倍、選択枝数が K 個のときの積分コストは 2^{K-2} 倍となる。つまり、積分区間の分割数を倍 n すると積分コストは n^{K-2} 倍となる。

一方、図-10に注目すると相対誤差は積分区間の分割数を増加させるにつれて小さくなり、真値に近づいている。つまり、数値積分の精度を十分に高くすることで、計算結果は真値に近づいていく。つまり、数値積分の精度を上げることで、選択確率確率の計算精度も上がるが、同時に計算にかかる時間も増加する。

Tridiagonal Decomposition Methodsは以下の式(18)によって選択確率を求めている。

$$g_{K-1}(y) = \Phi(z_{j,K-1} - c_{K-2}y)$$

$$g_i(y) = \int_{-\infty}^{z_{j,i} - c_{i-1}y} \phi(t) g_{i+1}(t) dt$$

$$\Phi(\mathbf{Z} < \mathbf{z}_j; \mathbf{T}_j) = g_1 = \int_{-\infty}^{z_{j1}} \phi(t) g_2(t) dt \quad (18)$$

Tridiagonal Decomposition Methodsは、Separation-of-Variables Methodsと同様に積分の中に積分が存在する構造であり、積分区間の分割数を倍 n すると積分コストは n^{K-2} 倍となる。

一方、図-10に注目するとSeparation-of-Variables Methodsと同様に相対誤差は積分区間の分割数を増加させるにつれて小さくなり、真値に近づいている。特にこの手法は近似の精度が高いため、積分区間の分割数が20個以上のときには非常に真値に近い計算結果となった。以上より、Separation-of-Variables Methodsと同様に数値積分の精度を上げることで、選択確率率の計算精度も上がるが、同時に計算にかかる時間も増加する。

Spherical Radial Transformation Methodsは以下の式(19)によって選択確率を求めている。

$$P = \int_{\|\mathbf{x}\|=1} \frac{2^{-\frac{K-1}{2}}}{\Gamma(\frac{K-1}{2})} \int_{r_1(\mathbf{x})}^{r_2(\mathbf{x})} r^{K-2} e^{-\frac{r^2}{2}} dr dU(\mathbf{x}) \quad (19)$$

ここで、合成シンプソン公式の積分区間の分割数を2倍にすると式(19)の積分コストは2倍に、分割数を3倍にすると式(19)の積分コストは3倍にといったように、等倍で積分コストが大きくなる。

一方、図-10に注目すると相対誤差は一定の値をとっていることがわかる。これは、Variable Reorderingと同様に数値積分の精度が計算結果にあまり影響を与えないということを示している。つまり、数値積分の精度を上げると計算にかかる時間が増加するが、選択確率はある一定の値で収束しそれ以上真値に近づくことはない。

(4) 仮想ネットワークにおける確率計算結果のまとめ

Variable Reorderingは他の3つの手法と比較すると誤差が大きくなった。特に、他の3つの手法と違い相関の影響を受けやすく、相関が中程度あるときにはおよそ10%の誤差が生じてしまった。また、選択枝数の増加に伴い誤差も大きくなった。ただし、計算速度は4つの手法の中で1番速いため、計算速度を重視する際には適している。

Separation-of-Variables MethodsとTridiagonal Decomposition Methodsは効用差の変化による影響も相関の違いによる影響もほとんど受けることなく、常に高い精度を維持していた。特に、Tridiagonal Decomposition Methodsの精度は非常に高く、選択枝が4つであれば相対誤差は常に0.1%未満であった。選択枝数が大きい場合にはSeparation-of-Variables Methodsの誤差も大きくなるがその増加量はVariable Reorderingより小さい。したがって、計算の精度を重視する際にはこれらの手法が適している。

Spherical Radial Transformation MethodsもSeparation-of-Variables MethodsとTridiagonal Decomposition Methodsと同様に効用差、相関の影響がほとんど見られなかった。ただし、計算精度は2つの手法と比較すると少し精度が低くなった。

以上より、4つの手法の間には計算速度が速い手法は計算精度が低く、計算精度が高い手法は計算速度が遅く、計算精度と速度のトレードオフ関係が見られる。また、数値積分の精度と計算速度の間にもトレードオフ関係は存在するため、計算精度と速度を考慮して適切な設定をする必要があることが確認された。

4. 実ネットワークの旅客配分計算への適用

実際の首都圏鉄道ネットワークの旅客配分計算に解析的積分近似手法を適用した際の分析を行う。実ネットワークでの分析を行うに当たって、重複経路の長さに応じて共分散構造が変化する構造化プロビットモデル²⁰⁾を用いる。

今、効用の誤差項を経路長に依存する誤差 e^1 と経路固有の誤差 e^0 の互いに独立な2つの項に分離して考える。このとき経路 k の誤差項は式(20)のように表現することができる。

$$e_k = e_k^1 + e_k^0 \quad (20)$$

ここで、経路長に依存する誤差 e_k^1 は経路線長 L_k に比例して算出されると仮定する。一方、経路固有の誤差 e_k^0 は経路ごとに独立に発生すると仮定する。単位長さ当たり発生する誤差分散を σ_1^2 、経路固有の誤差分散を σ_0^2 と置き、分散比 η を用いると、分散共分散行列は以下のように表現される。

$$\Sigma = \sigma_0^2 \begin{bmatrix} \eta L_1 + 1 & \eta L_{12} & \cdots & \eta L_{1K} \\ \eta L_{12} & \eta L_2 + 1 & \cdots & \eta L_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \eta L_{1K} & \eta L_{2K} & \cdots & \eta L_K + 1 \end{bmatrix}$$

$$\eta = \frac{\sigma_1^2}{\sigma_0^2}$$

本研究で利用するデータは、平成22年度大都市交通センサス（首都圏）のうち、選択可能な経路が3つ以上存在する場合のODペアデータである。全ODペア数は799,718であり、選択可能な経路の数の平均は9.9個である。配分計算に用いる効用関数パラメータも平成22年度大都市交通センサスで推定されたものを使用する。

ここでは、4手法の中で計算速度が1番速いVariable Reorderingを用いて配分計算を行う。同時に乱数生成数を10,000回行った際のGHKシミュレーターによる配分計算を行い、比較を行い、考察する。

図-11にVariable Reorderingを用いた際の確率計算結果と真値との絶対誤差を各選択枝ごとに示す。ここでいう真値とは乱数生成数を10,000回行った際のGHKシミュレーターを用いた際の確率計算結果のことである。

まず、計算精度に注目する。図-11を見ると選択枝数の増加に伴い、絶対誤差の平均値、最大値、中央値のすべてが大きくなっていることがわかる。これは、選択枝数の増加に伴い近似計算の精度が低くなることを示唆している。また、選択枝数の増加に伴い四分位範囲も大きくなっており、絶対誤差のばらつきも大きくなっている。つまり、選択枝数が大きい場合には近似計算の結果が不安定になる。

全体の計算結果に注目すると絶対誤差の平均値は1.82%であり、四分位範囲も0.36% から2.58%であった。さらに、絶対誤差の値が5%以下となるようなODペアの割合は全体の92.37%と大部分は十分な精度で計算することができる。しかし、選択枝が3つの場合でも絶対誤差の最大値は9.85%であり、選択枝が14個の場合には最大値が19.1%と高く、選択枝数にかかわらず近似の精度が低くなる場合があることに注意しなければならない。

では、どのような場合に近似の精度が低くなるのだろうか。以下ではどのような場合に近似の精度が低くなるのかを選択枝が3つの場合を例に考察する。まず、選択枝が3つのODペアのうち、誤差が最大となったケースと誤差が非常に小さいケースを抽出し、それらの経路の選択確率を図-12、図-13に示す。また、そのときの重複距離は表-1、表-2に示す。

表-1を見ると誤差が最大となったケースでは、経路1と経路2の重複距離が各経路の総距離のおよそ60%、経路2と経路3もおよそ60%、経路3と経路1はおよそ30%である。これは仮想ネットワークでVariable Reorderingを用いた選択確率計算の検証をした際に相関が中程度ある場合に選択確率が真値より大きくなった結果と一致する。一方、表-2を見ると経路1と経路2と経路3のすべての経路でほぼ等しい経路を通り、経路3のみが他の2つの経路より総距離が長くなる。つまり、相関が強いネットワークでは、誤差が小さくなる。同様に相関が弱い場合にも誤差が小さくなると考えられる。以上より、実際のネットワークに解析的積分近似計算に基づくプロビット型経路選択モデルを適用した際にも、相関の影響を強く受けることが示唆される。

最後に、計算速度について考察する。Rでの計算時間(16GB, Intel Core i7 @ 2.80Ghz)はVariable Reorderingを用いたときが7.9時間、GHKシミュレーターを用いたときが42.4時間であり、5分の1以下に短縮することができた。計算時間は、プログラムコードの改善による更なる短縮が期待できる。また、数値積分を行う際の積分区間の分割数を増加させると計算精度は高くなるが、計算速度は遅くなる。逆に、積分区間の分割数を減少させると計算速度が速くなり、計算精度が低くなる。そのため、計算精度要件を満たす適切な設定を見極めなければならない。

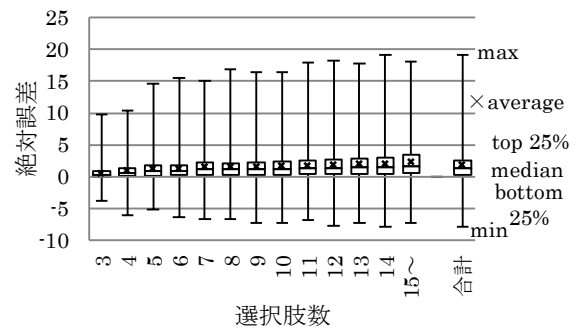


図-11 各選択枝ごとの選択確率計算結果のbox plot

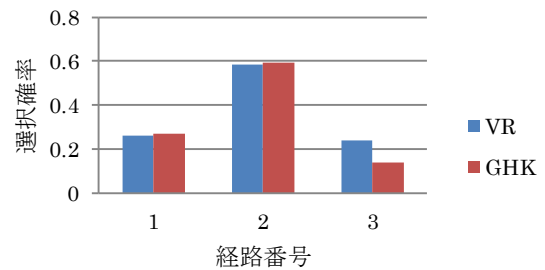


図-12 誤差が大きいケースの選択確率

表-1 誤差が大きいケースの重複距離

経路番号	1	2	3
1	72.5km	43.6km	22.8km
2	43.6km	68.1km	45.1km
3	22.8km	45.1km	74.1km

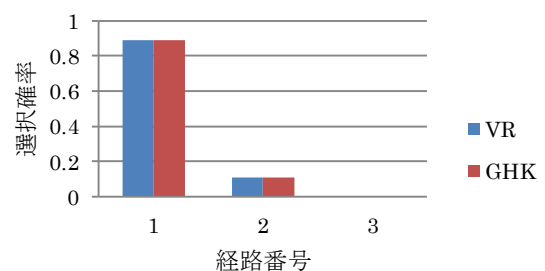


図-13 誤差が小さいケースの選択確率

表-2 誤差が小さいケースの重複距離

経路番号	1	2	3
1	21.5km	21.5km	21.5km
2	21.5km	21.9km	21.5km
3	21.5km	21.5km	41km

5. おわりに

本研究は、4つの解析的な積分近似手法をプロビットモデルに適用し、各手法の特徴の把握を行い、それらの有用性の検討を目的としたものである。

まず、4つの解析的な積分近似手法の特徴をまとめ、各手法の計算精度と速度の関係を示した。

次に、4つの手法を簡易ネットワークを対象にプロビット型経路選択モデルに適用し、選択確率の精度検証を行った。簡易ネットワークの効用差が変化した場合、相関が変化した場合、選択枝数が変化した場合を想定した精度検証を行った。選択確率の計算結果は、選択枝数が少ない場合は4つの手法のいずれも高い精度で近似することができ、その有用性が明らかとなった。ただし、Variable Reorderingは選択枝が少ない場合でも相関の変化によってプロビット確率を真値よりも大きく評価してしまうことがある。選択枝を増加させた際には、Variable ReorderingとSeparation-of-Variables Methodsのいずれも選択枝の増加に伴い、誤差は大きくなった。また、数値積分の精度と計算速度の間にはトレードオフ関係があるためプロビットモデルによる経路選択モデルを分析する際は、計算精度と速度を考慮して適切な設定を行う必要があることも確認された。

さらに、実際の東京圏鉄道ネットワークに解析的積分近似手法を適用した際は、大部分が十分な精度で近似することができた、しかし、選択枝数が大きい場合と経路間の相関が中程度の場合には計算精度が低くなることに注意しなければならない。また、計算速度についてはVariable Reorderingを用いた配分計算がGHKシミュレーターを用いた配分計算よりも速くなった。計算速度に関してはプログラムコードの改善による更なる時間短縮が期待できる。

本研究では、選択枝数が増加した場合のプログラムがVariable ReorderingとSeparation-of-Variables Methodsの2つの手法しか作成できなかったため、他の2つの手法は選択枝が4つの場合しか検証できなかった。Tridiagonal Decomposition Methodsは選択枝が4つの場合には非常に高い精度で近似することができたため、選択枝数が増加した際も、高い精度が期待できる。今後、本研究で作成できなかった2つの手法のプログラムを作成し精度を検証することで選択枝数が増加した際の有用性も把握できると考えられる。

また、解析的な積分近似を用いる最大の利点は計算の高速化である。計算速度に関してはより高度なプログラムコードの開発による更なる速度向上の余地が大いにあり得ると考えられる。

謝辞：本研究で使用した首都圏鉄道ネットワークデータ

は、社会システムの山下良久様よりご提供いただいたものです。ここに記して感謝の意を表します。

参考文献

- 1) 屋井鉄雄, 中川隆広, 石塚順一: シミュレーション法による構造化プロビットモデルの推定特性, 土木学会論文集, No.604, pp.11-21, 1998.
- 2) Clark, C.E.: The greatest of a finite set of random variables, *Operation Research*, Vol.9, pp.145-162, 1961.
- 3) Geweke, J.: Bayesian inference in econometric models using Monte Carlo integration, *Econometrica: Journal of the Econometric Society*, Vol.57 pp.1317-1339, 1989.
- 4) Hajivassiliou, V.A. and McFadden, D.L.: The method of simulated scores for the estimation of LDV models, *Econometrica*, Vol.66, No.4, pp.863-896, 1998.
- 5) Keane, M.P. and Runkle, D.E.: Testing the rationality of price forecasts: New evidence from panel data, *The American Economic Review*, Vol.80, No.4, pp.714-735, 1990.
- 6) Keane, M.P.: Simulation estimation for panel data models with limited dependent variables, *Handbook of Statistics*, Vol.11, pp.545-571, 1993.
- 7) Genz, A. and Bretz, F.: *Computation of Multivariate Normal and t Probabilities*, Springer, 2009.
- 8) 国土交通省: 運輸政策審議会 (<http://www.mlit.go.jp/singikai/unyusingikai/unseisin/unseisin.html>), 2000.
- 9) Schervish, M.J.: Algorithm AS 195: Multivariate normal probabilities with error bound, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol.33, No.1, pp.81-94, 1984.
- 10) Gibson, G.J. and Glasbey, C.A. and Elston, D.A.: Monte Carlo evaluation of multivariate normal integrals and sensitivity to variate ordering, *Advances in Numerical Methods and Applications, Sofia*, pp.120-126, 1994.
- 11) Geweke, J.: Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities, : *Computing science and statistics: Proceedings of the 23rd symposium on the interface, Citeseer*, Vol.23, pp.571-578, 1991.
- 12) Genz, A.: Numerical computation of multivariate normal probabilities, *Journal of Computational and Graphical Statistics*, Vol.1, No.2, pp.141-149, 1997
- 13) Schläfli, L.: On the multiple integral $\int^n dx_1 \dots dx_n$, whose limits are $p_1 = a_1x_1 + b_1y_1 + \dots + h_1z_1 > 0, p_2 > 0, \dots, p_n > 0$, and $x_2^2 + y_2^2 + \dots + z_2^2 < 1$, *Quart. J. Pure Appl. Math.*, Vol.2, pp.69-301, 1858.
- 14) Miwa, T. and Hayter, A. and Kuriki, S.: The evaluation of general non-centred orthant probabilities, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol.65, No.1, pp.223-234, 2003.
- 15) Craig, P.: A new reconstruction of multivariate normal orthant probabilities, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol.70, pp.227-243, 2008.
- 16) Deak I: Three digit accurate multiple normal probabilities, *Numerische Mathematik*, Vol.35, pp.369-380, 1980.
- 17) Richter W.D.: Geometric approach to the Gaussian law. In: Mammitzsch, V, Schneeweiß, H, (eds.) *Symposia Gaussiana, Walter de Gruyter*, Berlin, pp.25-46, 1994.
- 18) Monahan J. and Genz A.: Spherical-radial integration rules

for Bayesian computation. *Journal of the American Statistical Association*, Vol.92 pp.664-674, 1997

19) Somerville P.N.: Multiple testing and simultaneous confidence intervals: Calculation of constants, *Computational*

Statistics and Data Analysis, Vol.25, pp.217-233, 1997.

20) 屋井鉄雄, 岩倉成志, 伊藤誠: 鉄道ネットワークの需要と余剰の推計法について, *土木計画学研究・論文集*, No.11, pp.81-88, 1998.

MULTINOMIAL PROBIT MODEL WITH ANALYTICAL APPROXIMATION FOR RAIL-ROUTE CHOICE ANALYSIS

Junpei KAMEGAI and Daisuke FUKUDA

Since multinomial probit model (MNP) can appropriately capture the correlations among overlapped routes in the highly dense network, it is suitable for rail-route choice model in metropolitan area. However, multidimensional integral included in MNP cannot be computed analytically. Therefore approximation with Monte Carlo Simulation has played a pivotal role. Towards the enhancement of demand forecasting, it would become more important to consider taste variations of travelers. The speed up of computing MNP model would be helpful for practitioners. This study examines MNP with four different analytical approximations for rail-route choice applications and demonstrates to compute route choice probabilities in rail-route assignment in network of Tokyo Metropolitan Area so that it is compared with the previous method. We find that there is availability of analytical approximation.