様　式　C－19、F－19－1、Z－19　(共通)

科研費 KAKENHI

科学研究費助成事業　　研究成果報告書

平成 29 年　 6 月　 5 日現在

機関番号：１４５０１
研究種目：若手研究(B)
研究期間：2014～2016
課題番号：２６８２０２１７

研究課題名（和文）空間統計学に基づく欠損データの復元のための新たな方法論の開発と実用化

研究課題名（英文）Development of imputation methods for spatially dependent missing data

研究代表者
瀬谷　創（SEYA, HAJIME）

神戸大学・工学研究科・准教授

研究者番号：２０５８４２９６

交付決定額（研究期間全体）：（直接経費）　　2,400,000円

研究成果の概要（和文）：実施者らの既往研究では，データの空間相関を考慮した欠損データの補完方法を提案した．しかし，そこでは，被説明変数yの空間相関のみが考慮され，説明変数Xの欠損復元においては，変数間の相関情報のみが利用されていた．そこで本課題では，Xの復元においても，空間情報を利用する形のモデル開発を行った．その結果次の2点が明らかになった．[1]空間情報を用いることで，正の空間相関の強い連続変数（例えば，賃料，アクセシビリティ）については，補完正確度を大きく向させられる可能性がある．[2]局所的な変動が大きい変数，例えば床面積や築年数については，空間情報の導入が逆にRMSEやバイアスの悪化を引き起こす場合が多い．

研究成果の概要（英文）：We developed new imputation method for spatially dependent missing data. The method attempts to apply spatial information to the prior distribution of missing values. We found that [1] By using spatial information for prior distribution, the imputation accuracy of some spatially auto-correlated explanatory variables can be improved (e.g., rent price, accessibility index). [2] Prior spatial information on missing data may not be effective for variables with a high {nugget/sill} ratio; in other words, the measurement error and/or micro-scale variation is dominant in total variation (e.g., FAR(Ratio) and age).

研究分野：土木計画学

キーワード：地理情報　空間相関　欠損データ

１．研究開始当初の背景

データ欠損は，分野を問わず大きな課題である．例えば，不動産データでは，マンションが有する設備の一部は非公開である場合があるし，経済調査においては，回答項目が多岐にわたることもあり，記入漏れや記入誤りが生じやすい．このような欠損データへの対処法として，最も標準的に用いられている手法は，欠損を含むデータは捨て去るという，リストワイズ除去と呼ばれる手法である．King et al. (2001) の調査によれば，93～97 年に彼らが調査した政策科学系の雑誌では，94%の研究で同手法による対応が行われていたと指摘されている．しかし，リストワイズ除去が有効なのは, MCAR（missing completely at random）が成り立つ場合，すなわち欠損するかどうかが変数自身や他の変数に依存せず，完全にランダムであるという，極めて限定的なケースのみであり，MCAR が成り立たない場合，モデルのパラメータ推定値に深刻なバイアスが発生する可能性がある．

これに対して，統計学の分野では，多重代入法と呼ばれる有効な代替案が提案されており，わが国でも経済調査の欠損値補完に用いるような動きも出てきている．実施者らは，堤ら (2008)，Seya and Tsutsumi (2011) において，ベイズ統計学をベースとした，回帰モデルに基づく多重代入法のアルゴリズムと位置づけられる Knight et al.(1998) の手法の, 空間統計モデルへの拡張を行った（以下，空間 MCMC 法[マルコフ連鎖モンテカルロ]と称する）．土木計画分野で用いられることの多いデータ，例えば人口データや不動産データには，「距離の近いデータが似たような傾向を示す」という空間的自己相関（以下，簡単に空間相関と称す）と呼ばれる性質があり，実施者らの提案した空間 MCMC 法では，この特性が明示的に考慮されている．その結果，
A）リストワイズ除去との比較で，パラメータ推定値のバイアスの縮小と精度（precision）の向上
B）Knight et al.(1998) との比較で，被説明変数 y の欠損補完精度の大幅な向上（約 25%）
が達成できることを示した．計量地理学の第一人者であるケンブリッジ大学の Robert Haining は，　著書 Haining (2003) において，“No methods appear to have been developed in the spatial literature to deal with missing values in the explanatory variables or the explanatory and the response variables.” と指摘しており，この点において，被説明変数 y と説明変数 X の欠損どちらにも対応できる申請者らの空間 MCMC 法は新規性を持つと考えられる．また，空間 MCMC 法は，回帰モデルをベースとしているため，パラメータ推定とクリギングと呼ばれる任意点におけるデータの内挿（予測）を，欠損値の不確実性を考慮しながら同一フレームで実行できるという点も大きな利点である．

参考文献

- 堤盛人，吉田靖，瀬谷創，川口有一郎 (2008) 「MCMC 法によるデータ欠損問題と空間的相関を考慮した不動産賃料予測モデル：東京 23 区における賃貸マンション市場の実証分析」，『ジャレフ・ジャーナル』，3，1‑13.
- Haining, R. (2003) Spatial Data Analysis, Cambridge University Press, Cambridge.
- King, G., Honaker, J., Joseph, A. and Scheve, K. (2001) Analyzing incomplete political science data: An alternative algorithm for multiple imputation, American Political Science Review, 95 (1), 49–70.
- Knight, J.R., Sirmans, C.F., Gelfand, A.E. and Ghosh, S.K. (1998) Analyzing real estate data problems using the Gibbs sampler, Real Estate Economics, 26 (3), 469¬–492.
- Seya, H. and Tsutsumi, M. (2010) Spatial hierarchal Bayesian model for missing data imputation, presented at the Fourth World Conference of the Spatial Econometrics Association, Chicago, 9–12 June.

２．研究の目的

一方で，空間 MCMC 法は，いくつかの改善すべき課題を持っていた. 本課題では,「空間情報を欠損補完に利用する」という視点からモデルの改良を行うことを目的に実施したものである．

３．研究の方法

堤ら (2008)，Seya and Tsutsumi (2011) では，被説明変数 y の空間相関のみが考慮され，説明変数 X の欠損復元においては, 既往の多くの手法同様，変数間の相関情報のみを利用していた．これは，地点 i におけるある説明変数 x1i が欠損しているとき，x1i を，別の変数 x2i との相関情報を用いて復元するアプローチである．しかし，欠損値にも，「距離の近い欠損値には似たような傾向がある」という空間相関が存在すると考えられ，これを明示的に考慮することで欠損値補完の精度を向上させられる場合も多いと考えられる．そこで本課題では，観測された説明変数の空間内挿（クリギング）によって得られる予測分布を欠損説明変数の事前情報として利用することで, この点の改良を試みた. 具体的には，欠損説明変数の復元において，事前情報を
[A] 用いない場合（事前分布が定数）
[B] 観測標本の平均と分散に基づく正規分布で与える場合
[C] クリギングで与える場合
を比較し，どのような状況下で空間情報が有用になりうるか，実際の不動産データを用いて，定量的に明らかすることを目的とした．通常のベイズ統計分析では，主観性を排除するために，可能な限り無情報な事前分布を用

いることが多い．しかし，欠損補完においては，利用できる情報は積極的に利用して精度を上げることが重要であると考えられ，本課題ではこの点についての検証を行ったものである．

４．研究成果

得られた主要な結果は，次の 2 つに集約できる．

[1] 空間情報を用いることで，正の空間的自己相関の強い連続変数（例えば，賃料，アクセシビリティ指標など）については，補完正確度を大きく向上させられる可能性がある．

[2] ナゲット/シル比（局所的な変動や測定誤差）が大きい変数，例えば床面積や築年数などについては，空間情報の導入が欠損復元における RMSE や回帰係数のバイアスの悪化を引き起こす場合が多く，事前情報を導入せず，データの相関情報のみから復元を試みることが有用な可能性がある．

以下では，開発したモデルの具体的なアルゴリズムについて説明する．

本課題では，次式に示される地球統計モデルをベースモデルとして用いた．

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \boldsymbol{\Sigma})\,; \qquad (1)$$

ここで，y は $n\times 1$ の従属変数ベクトル，$\boldsymbol{X}$ は $n\times k$ の説明変数行列，β は $k\times 1$ のパラメータベクトルである．分散共分散行列は，$\Sigma = \tau^2 I + \sigma^2 \boldsymbol{H}(\phi)$，と定式化する．ここで，$\boldsymbol{I}$ は $n\times n$ の単位行列，$\boldsymbol{H}$ はその i, j 要素を相関関数で与える $n\times n$ の相関行列，τ^2，σ^2，ϕ はそれぞれナゲット（nugget），パーチャル・シル（partial-sill），レンジ（range）と呼ばれるパラメータである．ナゲットとパーシャル・シルの和は，シル（sill）と呼ばれる．Σ の{i , j}要素：Σ_{ij} は，例えば球型（spherical）の共分散関数 C(d_{ij})を用いた場合

$$C(d_{ij}) = \begin{cases} 0 & if\ \ d_{ij} > \phi \\ \sigma^2\left[1 - \frac{3}{2}\frac{d_{ij}}{\phi} + \frac{1}{2}\left(\frac{d_{ij}}{\phi}\right)^3\right]\cdot & if\ \ 0 < d_{ij} \le \phi \\ \tau^2 + \sigma^2 & otherwise \end{cases} \qquad (2)$$

で与えられる．ここで，d_{ij} は，地点 i, j 間のユークリッド距離を示す（ただし，等方性を仮定）．このように，地球統計モデルでは，分散共分散行列を距離の関数で直接構造化するアプローチをとる．

今，共分散関数のパラメータをベイズ推定するにあたり，パラメータ集合 $\theta = (\beta', \sigma^2, \tau^2, \phi)'$ に関する事前分布は独立，すなわち $p(\theta) = p(\beta)p(\sigma^2)p(\tau^2)p(\phi)$ を満たすとしよう．ここで，ベイズの定理 $p(\theta|\boldsymbol{y}) \propto f(\boldsymbol{y}|\theta)p(\theta)$ より，事前確率密度関数を尤度関数 f と合成することで各パラメータの事後確率密度関数を求めることができる．ここで，計算の都合上，式 (1) を，次のような条件付きモデルに書き換える．

$$\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\eta},\tau^2 \sim N(\boldsymbol{X\beta}+\boldsymbol{\eta},\tau^2\boldsymbol{I}); \qquad (3)$$

$$\boldsymbol{\eta}|\sigma^2,\phi \sim N(\boldsymbol{0},\sigma^2\boldsymbol{H}(\phi));$$

ここで，$\boldsymbol{\eta}$ は空間ランダム効果 $\eta(s_i)$ からなる $n\times 1$ のベクトルである．このような階層ベイズモデルとして定式化することで，分散パラメータに関する条件付き事後分布が簡単な形で書けるようになり，MCMC 計算が簡便になる．

事後確率密度関数は，次式のような事前確率密度関数と尤度の積に比例する．

$$f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\eta},\tau^2) f(\boldsymbol{\eta}|\sigma^2,\phi) p(\boldsymbol{\beta}) p(\tau^2) p(\sigma^2) p(\phi)\,; \qquad (4)$$

ここで，各パラメータの事前分布を，$\boldsymbol{\beta} \sim N(\dot{\boldsymbol{\beta}}, \dot{E})$，$\sigma^2 \sim IG(a_\sigma/2,\ b_\sigma/2)$，$\tau^2 \sim IG(a_\tau/2,\ b_\tau/2)$，$\phi \sim IG(a_\phi/2,\ b_\phi/2)$ と設定する．ただし，$IG(a,b)$ は，a を形状，b を尺度の超パラメータとする逆ガンマ分布を示す．各超パラメータは，$\dot{\boldsymbol{\beta}} = \boldsymbol{0}$，$\dot{E} = 10^5 \times \boldsymbol{I}$，$a_\sigma/2 = a_\sigma/2 = a_\phi/2 = 2.1$，$b_\sigma/2 = b_\tau/2 = 0.55$，$b_\phi/2 = 16.5$ と設定した．これらの事前分布と式 (3) より，各パラメータに関する条件付事後分布が，次式のように求められる．

$$\boldsymbol{\beta}\,|\,\boldsymbol{\eta},\tau^2,\boldsymbol{y},\boldsymbol{X} \sim N(\ddot{\boldsymbol{\beta}},\ddot{\boldsymbol{E}})\,; \qquad (5)$$

$$\boldsymbol{\eta}\,|\,\boldsymbol{\beta},\tau^2,\sigma^2,\phi,\boldsymbol{y} \sim N(\ddot{\boldsymbol{\eta}},\ddot{\boldsymbol{F}})\,; \qquad (6)$$

$$\tau^2\,|\,\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{y} \sim IG\left(a_\tau + n,\ b_\tau + (\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{\eta})'(\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{\eta})\right); \qquad (7)$$

$$\sigma^2\,|\,\boldsymbol{\eta},\phi \sim IG\left(a_\sigma + n, b_\sigma + \boldsymbol{\eta}'\boldsymbol{H}^{-1}(\phi)\boldsymbol{\eta}\right); \qquad (8)$$

$$p(\phi\,|\,\boldsymbol{\eta},\sigma^2) \propto \frac{p(\phi)}{|\boldsymbol{H}|^{1/2}} \times \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\eta}'\boldsymbol{H}^{-1}(\phi)\boldsymbol{\eta}\right); \qquad (9)$$

ただし，

$$\ddot{\boldsymbol{\beta}} = \left(\frac{1}{\tau^2}\boldsymbol{X}'\boldsymbol{X} + \dot{\boldsymbol{E}}^{-1}\right)^{-1}\left(\frac{1}{\tau^2}\boldsymbol{X}'(\boldsymbol{y}-\boldsymbol{\eta}) + \dot{\boldsymbol{E}}^{-1}\dot{\boldsymbol{\beta}}\right);$$

$$\ddot{\boldsymbol{E}} = \left(\frac{1}{\tau^2}\boldsymbol{X}'\boldsymbol{X} + \dot{\boldsymbol{E}}^{-1}\right)^{-1};$$

$$\ddot{\boldsymbol{\eta}} = \left(\frac{1}{\tau^2}\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{H}^{-1}(\phi)\right)^{-1}\left(\frac{1}{\tau^2}(\boldsymbol{y}-\boldsymbol{X\beta})\right);$$

$$\ddot{\boldsymbol{F}} = \left(\frac{1}{\tau^2}\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{H}^{-1}(\phi)\right)^{-1}.$$

条件付き分布の式から分かるように，レンジ

パラメータϕ以外は，条件付事後分布が標準的な分布となっており，容易に乱数を発生できる（ギブズ・サンプラー）．しかしながら，ϕに関する事後分布は標準的な分布とはなっておらず，MH アルゴリズムなどを用いる必要がある．本研究では，MH アルゴリズムのうち，実装が比較的容易な酔歩過程を用いることとする．具体的には，tを MCMC シミュレーションのステップ番号としたとき，

$$\ln(\phi)^{*} = \ln(\phi)^{(t-1)} + \upsilon_{t,}\ \upsilon_t \sim N(0,\zeta^2); \quad (10)$$

からϕの候補を発生する．採択確率は，

$$\alpha\left(\phi^{(r-1)},\phi^{*}\right) = \min\left(\frac{p(\phi^{*}\mid\eta,\sigma^2)}{p(\phi^{(t-1)}\mid\eta,\sigma^2)},1\right); \quad (11)$$

で与えられる． パラメータζ^2は，採択確率がおおよそ 40~50%程度となるように，試行錯誤で決定した．

ここまで，地球統計モデルの完全データ（full data）ケースにおけるパラメータのベイズ推定方法について説明してきた．以下では，これを欠損データを含む場合に拡張する．Knight et al. (1998) は，上述したようなベイジアン MCMC 法の枠組みの中で，欠損復元を行う方法を提案した．この方法は，imputation-posterior (IP) 法と呼ばれる多重代入法実行の代表的アルゴリズムのひとつであるが，Knight et al. (1998) の方法の特徴的な点は，回帰モデルの枠組みで$\boldsymbol{y}$と$\boldsymbol{X}$を明示的に分離し，両者の欠損復元を可能にしている点である．以下ではまず，この方法を説明し，次項においてそれを地球統計モデルに拡張した Seya and Tsutsumi (2011) の手法を説明する．なお，Seya and Tsutsumi (2011) はカテゴリカル変数の欠損に対応しておらず，また，定式化自体も MCMC 計算の効率性の観点から本課題で改良を試みた．今，次のような簡単な線形回帰モデルを考えよう．

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_0 + \varepsilon_i, \quad \varepsilon_i \sim N(0,\xi^2). \quad (12)$$

まず，欠損データが無いものと仮定しよう．このとき尤度関数は，$f(\boldsymbol{y}\mid\boldsymbol{\omega}) = (\xi^2)^{-n/2}\exp[-\sum(y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_0)/2\xi^2]$ で与えられる．ここで，$\boldsymbol{\omega} = (\beta_1,\beta_2,\beta_0,\xi^2)'$はパラメータ集合である．ベイズの定理より，$\boldsymbol{\omega}$への事前確率密度関数と尤度の積として，事後確率密度関数が$p(\boldsymbol{\omega}\mid\boldsymbol{y}) \propto f(\boldsymbol{y}\mid\boldsymbol{\omega})p(\boldsymbol{\omega})$と得られる．今，事前確率密度が$p(\omega) = p(\beta_1, \beta_2, \beta_0)p(\xi^2)$で与えられ，$p(\beta_1, \beta_2, \beta_0) = 1$（定数），$\xi^2 \sim IG(a_\xi, b_\xi)$を満たすと仮定する．このとき，各パラメータの条件付き事後分布は，次式により与えられる（**rest** は，注目する変数以外の変数からなるベクトルを示す）．

$$\beta_0 \mid \boldsymbol{rest} \sim N\left(\frac{\sum(y_i - \beta_1 x_{1i} - \beta_2 x_{2i})}{n},\frac{\xi^2}{n}\right); \quad (13)$$

$$\beta_1 \mid \boldsymbol{rest} \sim N\left(\sum x_{1i}(y_i - \beta_2 x_{2i} - \beta_0),\frac{\xi^2}{\sum x_{1i}^2}\right);$$

$$\beta_2 \mid \boldsymbol{rest} \sim N\left(\sum x_{2i}(y_i - \beta_1 x_{1i} - \beta_0),\frac{\xi^2}{\sum x_{2i}^2}\right);$$

$$\xi^2 \mid \boldsymbol{rest} \sim IG\left(a_\xi + \frac{n}{2}, b_\xi + \frac{1}{2}\sum(y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_0)^2\right). \quad (14)$$

これらの条件付き分布より，ギブズ・サンプラーを用いて標本を発生させれば，事後分布の評価が可能になる．

次に，欠損がある場合について考える．今，y_1 と x_{12} が両方欠損していると考えよう．ここでは，これらは連続変数とする．このとき，これらも「推定すべきパラメータ」と見れば，パラメータ集合が，$\boldsymbol{\omega}$から$\tilde{\boldsymbol{\omega}} = (\boldsymbol{\beta}',\xi^2,y_1,x_{12})'$，へと拡大され（augmented），欠損変数の条件付き事後分布が，

$$y_1 \mid \boldsymbol{\beta},\xi^2,x_{11},x_{21} \sim N\left(\beta_1 x_{11} + \beta_2 x_{21} + \beta_0,\xi^2\right); (15)$$

$$x_{12} \mid \boldsymbol{\beta},\xi^2,y_2,x_{22} \sim N\left(\frac{y_2 - \beta_2 x_{22} - \beta_0}{\beta_1},\frac{\xi^2}{\beta_1^2}\right); (16)$$

と与えられる（データ拡大法）．すなわち，欠損データの復元のためには，式 (15) と式 (16) をギブズ・サンプラーに追加するだけでよい．

さらに，前項で述べた手法を地球統計モデルに拡張する．今，式 (3) において，地点 i における不動産価格 y_i と，地点 j における l 番目の属性 x_{lj} の両者が欠損していると仮定する．このとき，パラメータ集合は，$\boldsymbol{\theta}$から$\tilde{\boldsymbol{\theta}}$

$=(\theta', y_i, x_{lj})'$ に拡大され，条件付き事後分布が次式のように得られる．

$$y_i \mid \boldsymbol{rest} \sim N\left(\boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{c}_i'\boldsymbol{\Sigma}_{-i-i}^{-1}(\boldsymbol{y}_{-i} - \boldsymbol{X}_{-i}\boldsymbol{\beta}), \upsilon_i^2\right);$$

$$x_{lj} \mid \boldsymbol{rest} \sim N\left(\frac{[y_j - \{\boldsymbol{x}_{-lj}'\boldsymbol{\beta}_{-l} + \boldsymbol{c}_j'\boldsymbol{\Sigma}_{-j-j}^{-1}(\boldsymbol{y}_{-j} - \boldsymbol{X}_{-j}\boldsymbol{\beta})\}]}{\beta_l}, \frac{\upsilon_j^2}{\beta_l^{\,2}}\right); \quad (17)$$

ここで， $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{k-1i}, 1)'$， $\boldsymbol{c}_i = \sigma^2\{\rho(\phi; d_{i1}), \ldots, \rho(\phi; d_{ii-1}), \rho(\phi; d_{ii+1}), \ldots, \rho(\phi; d_{in})\}'$， $\upsilon_i^2 = \tau^2 + \sigma^2 - \boldsymbol{c}_i'\Sigma_{-i-i}^{-1}\boldsymbol{c}_i$ であり， Σ_{-i-i} は， Σ から i 番目の行と列を除いたもの， $\boldsymbol{y}_{-i}$ と $\boldsymbol{X}_{-i}$ はそれぞれ $\boldsymbol{y}$ と $\boldsymbol{X}$ から i 番目の行を除いたもの， β_{-l} は， β から l 番目の行を除いたものであり， $\boldsymbol{x}_{-lj}$ は， $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{k-1j}, 1)'$ から l 番目の属性を除いたものである．

さて，以上の議論では，欠損データに対する事前分布は考えなかった．したがって，暗に事前分布として定数を仮定していたこととなる．これに対して，本課題では，標本平均 μ_l・分散 ψ_l^2 を用いて， $x_{lj} \sim N(\mu_l, \psi_l^2)$ という正規分布を仮定した．このとき，条件付き事後分布は，

$$x_{lj} \mid \boldsymbol{rest} \sim N\left(\frac{\psi_l^2\beta_l[y_j - \{\boldsymbol{x}_{-lj}'\boldsymbol{\beta}_{-l} + \boldsymbol{c}_j'\boldsymbol{\Sigma}_{-j-j}^{-1}(\boldsymbol{y}_{-j} - \boldsymbol{X}_{-j}\boldsymbol{\beta})\}] + \upsilon_j^2\mu_l}{\psi_l^2\beta_l^2 + \upsilon_j^2}, \frac{\psi_l^2\upsilon_j^2}{\psi_l^2\beta_l^2 + \upsilon_j^2}\right); \quad (18)$$

で与えられる．式 (18) を用いた場合，もし従属変数と l 番目の変数の相関関係が弱い，すなわち $\beta_l \sim 0$ の場合，事前分布の平均と分散が条件付き事後分布において支配的になる．逆に，従属変数と l 番目の変数の関係が強ければ，事前の平均と分散の寄与分は低くなる．したがって， $\beta_l \sim 0$ において，この定式化は式 (17) より安定性を持つと考えられる．しかしながら，このような事前情報を導入することで，説明変数の復元値が標本平均方向にスムージングされるため，欠損値や係数推定値にバイアスを生じさせる可能性もある．

そこで本研究では，事前分布として空間情報を用いることを試みる．これは，ある説明変数 x_{lj} が欠損している場合，それが標本平均に似ていると仮定するよりは，地理的に近接した地点 i における観測説明変数 x_{li} に似ていると考えるほうが自然であろうという考え(18)に基づく．

具体的には，次式のような通常クリギング予測値に基づく事前分布を用いる．

$$x_{lj} \sim N\left(\lambda_l + \boldsymbol{e}_{lj}'\boldsymbol{\Gamma}_{l-j-j}^{-1}(\boldsymbol{X}_{l-j} - \lambda_l\boldsymbol{1}_{-j}), v_{lj}^2\right); \quad (19)$$

ここで， λ_l は定数項パラメータ， $\boldsymbol{X}_{l-j}$ は l 番目の説明変数からなる $n \times 1$ のベクトル $\boldsymbol{X}_l$ から第 j 番目の成分を除いたもの， $\boldsymbol{1}_{-j}$ は 1 からなる $n \times 1$ ベクトルから j 番目の要素を除いたもの， $\boldsymbol{e}_{lj} = \sigma_l^2\{\rho(\phi_l; d_{j1}), \ldots, \rho(\phi_l; d_{jj-1}), \rho(\phi_l; d_{jj+1}), \ldots, \rho(\phi_l; d_{jn})\}'$， Γ_{l-j-j} は，説明変数 l に関する分散共分散行列 Γ_l から j 番目の行と列を除いたもの， $v_{lj}^2 \equiv \tau_l^2 + \sigma_l^2 - \boldsymbol{e}_{lj}'\boldsymbol{\Gamma}_{l-j-j}^{-1}\boldsymbol{e}_{lj}$， τ_l^2， σ_l^2， ϕ_l はそれぞれ説明変数 l に関するナゲット，パーチャル・シル，レンジである．

さて，今までの議論では， x_l が連続としてきたが，ここでは離散，すなわち $\eta_l^1, \ldots, \eta_l^Q$ のうちのいずれかの値をとると考えよう．このとき， x_{lj} の条件付き事後分布は， η_l^q に次の値を対応させ，

$$p_{lj}^q = \frac{\exp[-\{y_j - \boldsymbol{x}_{-lj}'\boldsymbol{\beta}_{-l} - \beta_l\eta_l^q - \boldsymbol{c}_j'\boldsymbol{\Sigma}_{-j-j}^{-1}(\boldsymbol{y}_{-j} - \boldsymbol{X}_{-j}\boldsymbol{\beta})\}^2/2\upsilon_j^2]}{\sum_{q'=1}^{Q}\exp[-\{y_j - \boldsymbol{x}_{-lj}'\boldsymbol{\beta}_{-l} - \beta_l\eta_l^{q'} - \boldsymbol{c}_j'\boldsymbol{\Sigma}_{-j-j}^{-1}(\boldsymbol{y}_{-j} - \boldsymbol{X}_{-j}\boldsymbol{\beta})\}^2/2\upsilon_j^2]}; \quad (19)$$

区間(0,1]を， $(0, p_{lj}^1], (p_{lj}^1 + p_{lj}^2], \ldots$ と区切り，(0,1] における一様分布から乱数を発生させて，それが $(\sum_{q'=0}^{q-1} p_{lj}^{q'}, \sum_{q'=0}^{q} p_{lj}^{q'}]$ に属したのならば， $x_{lj} = \eta_l^q$ を実現させればよい．また，式 (19) に事前分布を導入する場合，分子と分母の $exp(\bullet)$ を， $p(x_{lj} = \eta_l^q) \cdot \exp(\bullet)$ と，事前確率 $p(x_{lj} = \eta_l^q)$ に比例する形で置き換えればよい．

参考文献

- Knight, J.R., Sirmans, C.F., Gelfand, A.E. and Ghosh, S.K.: Analyzing real estate data problems using the Gibbs sampler, Real

Estate Economics, Vol.26, No.3, pp.469–492, 1998.

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 4 件）

① **山形与志樹・瀬谷創** (2015) **環境関連特許の空間・社会ネットワーク分析，応用地域学研究，** 15，67-85**.**

② Hajime Seya, Yoshiki Yamagata and Kumiko Nakamichi (2016) Creation of municipality level intensity data of electricity in Japan, Applied Energy, 162, 1336-1344.

③ Hajime Seya, Kumiko Nakamichi, Yoshiki Yamagata (2016) The residential parking rent price elasticity of car ownership in Japan, Transportation Research Part A: Policy and Practice, 85, 123-134.

④ Hajime Seya, Takahiro Yoshida, and Morito Tsutsumi (2016) Ex-post identification of geographical extent of benefited area by a transportation project: Functional data analysis method, Journal of Transport Geography, 55, 1-10.

〔学会発表〕（計 2 件）

① Hajime Seya, Yutaka Ito, Makoto Chikaraishi, and Makoto Tsukai (2015) Spatial competition among retail gasoline stations in Japan**，第** 29 **回** ARSC **研究発表大会，**2015 **年** 11 **月** 28～29 **日，慶應義塾大学.**

② **瀬谷創，丸田雅晴，力石真，藤原章正，張峻屹** (2016) **空間競争に関する離散選択モデ：ガソリンスタンドの撤退行動の実証分析，土木計画学研究・講演集，**Vol. 53 (CD-ROM)**.**

６．研究組織
(1)研究代表者
瀬谷創（SEYA, Hajime）
神戸大学・大学院工学研究科・准教授
研究者番号：20584296

(3)連携研究者
堤盛人（TSUTSUMI，Morito）
筑波大学・システム情報系社会工学域・教授
研究者番号：70292886