

# 空間統計モデルによる 新規出店計画店舗の売上予測

奥野 拓也

## 1. はじめに

厚生労働省が発表する平成28年度衛生行政報告例 [1] によると理容室は年々減少傾向にある反面、美容室は年々増加傾向にあり年間約3,000店舗が開業している。また店舗数と同様に美容師数も増加傾向にある。特に東京、神奈川、大阪、愛知、福岡では増加数全体の約80%を占めており大都市圏に集中している。この理由として大都市圏では女性の人口1,000人当たりの美容室数が東京(44位)、神奈川(47位)、大阪(40位)、愛知(43位)、福岡(38位)であり、相対的に不足していることがわかる。

その一方で美容業界の年間総売上高は減少を続けている。売上高の減少の理由として店舗あたりの顧客数の減少と客単価の低下(来店の長期サイクル化、低価格化など)が考えられる [2]。そのため事業を拡大するために店舗の新規出店計画は経営戦略上、重要な課題である。通常、新規店舗を開業するために立地選定のための市場調査を行う。立地調査とは候補地の商圈特性(住居者数、年齢層、性別、職業、競合店舗数など)の地域特性を調査し、予測集客数や店舗の集客要因をデータから特定していくことである。当然のことながら、立地調査には多額の予算が必要である。

立地選定は大まかに「地域の選定」後に「地点の選定」というプロセスを経て決定される。そのためマーケティング担当者らが過去の調査データや必要に応じて新たなデータを収集しながら地域の候補をいくつか選出していく必要がある。しかしながら費用と時間的コストの問題から関心をもつすべての土地の情報を必ずしも調査するわけではない。したがって、マーケティ

ング担当者らが関心をもつ土地における売上予測を知るためには、否応なしに既存店舗で蓄積された情報を利用した予測というプロセスが不可欠になる。そこで本研究では新規出店計画候補地を選出するために、これまで蓄積された複数店舗の購買履歴データと調査で得られる商圈の調査データからベイズモデルを用いて出店を支援するような情報抽出を目的とする。具体的には店舗の位置情報、POSデータ、オープンデータをモデルに組み込んで売上を予測し、その予測精度を実証的に評価する。またモデルに商圈の調査データを組み込むことで売上に寄与する要因を抽出することが可能となる。

## 2. 先行研究と本研究の位置づけ

市場予測の研究は古くから存在し、特に新製品を投入する際に対象とする製品の市場規模を計測・予測することによって、新製品の投入に関する意思決定を行うことが可能となることからさまざまな研究が発表されている [3, 4]。その代表的な統計モデルであるバスモデルは他人の影響を受けずに自らの意思で購入を決定するイノベータと、口コミなどから情報を収集しながら購入を決定するイミテータからなり製品の拡散過程を模擬するモデルである [5]。

製品の市場規模予測から店舗の売上予測に関する研究に目を向けると、売上予測を回帰によって直接的に予測するものや、来店人数を予測して客単価をかけることで店舗の売上を予測するものなどさまざまな方法が提案されている。たとえば山口ら [6] は売上に要因を分解するモデルを構築し、各店舗固有の情報に基づいて高精度な売上予測が可能となることを示した。また売上に要因を分解したことで、各要因を詳細に検討し背景にある合理的な理由を考察することを可能とした。

マーケティングにおける空間モデルを利用した新規出店店舗の候補地選定では Reilly [7] のモデルをはじめ多くの研究が報告されている。これらは地理的市場

おくの たくや  
NTT テクノクロス株式会社 IoT イノベーション事業部  
〒231-0032 神奈川県横浜市西区みなとみらい4-4-5  
okuno.takuya@po.ntt-tx.co.jp  
受付 18.7.25 採択 18.11.2

の物理的境界を決定する方法に重点を置き、小売商圏を決定する重力モデルが提案されている [8, 9].

空間を考慮した店舗の売上予測にはハフモデル [10] があり、実務でも用いられている。ハフモデルは購買地の選択確率を簡素な式で表現していることから、手元のデータを用いて来客数や商圏規模の推定に利用されるだけでなく、交通機関の発展と集客数の関係を解析的に調べるなど利用範囲が大きい。また Yang and Allenby [11] は居住地域の空間的な位置と嗜好の類似性を考慮したモデルを提案しているが、売上予測モデルではなく購買選択のモデルとして定式化している。これらの研究では空間的な情報を取り入れ、消費者の選択行動から店舗の売上を予測することが多い。しかし、本研究で提案する空間的な内挿を利用した売上予測モデルに関する研究は多くない。従来の研究では空間のネットワークを考え、その中で売上が最大となる地点を最適立地点としていた。そのため広いエリアにおける店舗立地選定には大規模ネットワークを考える必要があるため、計算時間に問題がある。一方、本研究で提案する空間内挿による手法では、ネットワークを構成する必要がないため出店候補地の位置情報が決まれば計算可能である。またモデルに回帰構造を組み込むことで、推定した回帰係数からハフモデルと同様にパラメータに関する議論が可能である。

### 3. モデル

#### 3.1 提案手法

本研究で提案する手法は、店舗の売上予測とその構成要因の推定に、クリギングを適用するものである。提案手法の特徴としては以下の点が挙げられる。

1. POS データと地域情報や調査データから空間相関を利用して新規出店店舗の売上高が予測可能となる。そのため、これまで候補選定に必要であった調査費用を軽減することが可能となる。
2. 売上高を構成する要因の影響度が推定可能となる。そのため、これまで進出していない地域における新規出店計画の予測が可能となる。

#### 3.2 クリギング

クリギングは D. G. Krige によって提案された空間内挿手法の一つで空間分布の予測のために体系化されてきた [12, 13]. クリギングでは距離の近い観測点のデータは大きな類似性をもつという空間相関構造を利用して、観測されたデータの共分散を空間距離の関数として構造化し、任意の地点に対して予測分散が最小となる空間内挿を行う手法である。

クリギングにはデータの空間自己相関のみを考慮する通常クリギング (ordinary kriging) や線形回帰モデルの誤差の共分散を共分散関数でモデル化する普遍型クリギング (universal kriging) などが存在する。その中でも空間的トレンドをモデルに取り込んだ普遍型クリギングは、距離の近い観測点のデータは大きな類似性をもつという空間データの性質を利用して空間現象の確率場に二次定常性を仮定し、観測されたデータの共分散を距離の関数として空間相関を構造化することによって任意地点での確率場の値を空間予測する手法である [14]. クリギングは埋没鉱量の空間補完のために導入されたが、空間情報を利用して地価を予測したり [15], センサーで NOx を観測してその広がりを予測するなどに応用されている [16]. これまでにクリギングを用いたさまざまな研究が報告されてきた。これは空間データのもつ性質として空間的相関が存在しているため、予測精度が高く、多くの研究で利用されている [17].

#### 3.3 モデルの定式化

地点  $s_j (j = 1, 2, \dots, N)$  における店舗の売上の対数を  $y(s_j)$  とする。本研究では  $y(s_j)$  が空間過程モデルで定義できると仮定し、 $\mathbf{y}(\mathbf{s}) = (y(s_1), \dots, y(s_N))^T$  を

$$\mathbf{y}(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \mathbf{w}(\mathbf{s}) + \boldsymbol{\epsilon}, \quad (1)$$

と定式化する。ここで  $\boldsymbol{\mu}(\mathbf{s})$  は確率変動を含まない地点  $\mathbf{s}$  に依存する関数であり、大域の変動を表す。残りの二つは確率の変動を含む項で一方は空間的な影響を考慮する  $\mathbf{w}(\mathbf{s})$  と非空間的な誤差項を表す  $\boldsymbol{\epsilon}$  である。 $\mathbf{w}(\mathbf{s})$  をガウシアンプロセスであると仮定することで普遍型クリギングとして定式化できる。本研究で用いる普遍型クリギングは Valente et al. [18] のモデルと同じ構造であるが、モデルの適用対象が異なる。

式 (1) を構築する各項を式 (2), (3), (4) と定式化する。

$$\boldsymbol{\mu}(\mathbf{s}) = \mathbf{x}^T \boldsymbol{\beta} \quad (2)$$

$$\mathbf{x} = \begin{bmatrix} 1 & x_{s_1,1} & \cdots & x_{s_1,K-1} \\ \vdots & \ddots & \ddots & \vdots \\ 1 & x_{s_N,1} & \cdots & x_{s_N,K-1} \end{bmatrix}$$

$$\mathbf{w}(\mathbf{s}) \sim MVN(\mathbf{0}, \sigma^2 \mathbf{C}(\mathbf{s}, \mathbf{s}'; r)) \quad (3)$$

$$\boldsymbol{\epsilon}(\mathbf{s}) \sim MVN(\mathbf{0}, \tau^2 \mathbf{I}) \quad (4)$$

ここで  $\mathbf{x}$  は  $N \times K$  の定数項を含む説明変数行列を表し、 $\boldsymbol{\beta}$  は  $K \times 1$  の回帰係数ベクトル、 $T$  は転置を表す。また  $\mathbf{C}(\mathbf{s}, \mathbf{s}'; r)$  はカーネル関数、 $\mathbf{I}$  は単位行列、

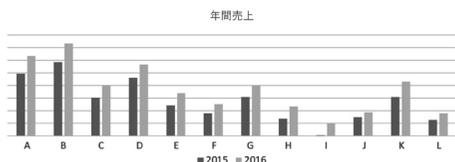


図1 各店舗の2015(左)、2016(右)年度の年間総売上。ただし店舗規模が特定されないように縦軸は非表示とした。

$r, \sigma, \tau$  はパラメータである。カーネル関数は店舗間の距離が近いほど影響が高く、遠いほど影響が小さくなることを表すため、地点間距離  $d_{i,j}$  を変数とした式(5)で表す。

$$C(s_i, s_j; r) = \exp\left(-\frac{d_{i,j}^2}{r^2}\right) \quad (5)$$

ここで  $r$  はパラメータ、 $d_{i,j}$  は地点  $s_i, s_j$  のユークリッド距離を表す。そのため分散共分散行列は  $C(s_i, s_j; r) = C(s_i, s_j; r)^T$  と対称行列となる。

$\theta = (\beta, r, \sigma, \tau)$  とし、尤度関数と事前分布はそれぞれ

$$\begin{aligned} \mathbf{y} | \theta &\sim N(\mathbf{x}^T \beta + \mathbf{w}, \tau^2 \mathbf{I}) \\ p(\theta) &= p(\beta)p(\sigma^2)p(\tau^2)p(r) \end{aligned} \quad (6)$$

と仮定する。したがって事後分布は

$$\begin{aligned} p(\theta | \mathbf{y}) &= N(\mathbf{y} | \mathbf{x}^T \beta + \mathbf{w}, \tau^2 \mathbf{I}) N(\mathbf{w} | 0, \sigma^2 \mathbf{C}) p(\beta, r, \sigma, \tau) \\ &= N(\mathbf{y} | \mathbf{x}^T \beta + \mathbf{w}, \tau^2 \mathbf{I}) N(\mathbf{w} | 0, \sigma^2 \mathbf{C}) \times \\ &\quad p(\beta)p(r)p(\sigma)p(\tau) \end{aligned}$$

と定式化できる。ただし実際の推定には尤度関数を  $\mathbf{w}$  で周辺化した、すなわち積分により  $\mathbf{w}$  を除去した

$$\begin{aligned} N(\mathbf{y} | \mathbf{x}^T \beta, \sigma^2 \mathbf{C} + \tau^2 \mathbf{I}) &= \frac{1}{(2\pi)^{N/2} |\sigma^2 \mathbf{C} + \tau^2 \mathbf{I}|^{1/2}} \times \\ &\quad \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{x}^T \beta)(\sigma^2 \mathbf{C} + \tau^2 \mathbf{I})(\mathbf{y} - \mathbf{x}^T \beta)\right] \end{aligned}$$

を利用する。

### 3.3.1 事前分布の設定

式(7)のように  $\beta, \sigma^2, \tau^2, r$  の事前分布を設定する。

$$\begin{aligned} \beta &\sim MVN(\mu_\beta, \Sigma_\beta) \\ \sigma^2 &\sim IG(\alpha_{\sigma^2}, \beta_{\sigma^2}) \\ \tau^2 &\sim IG(\alpha_{\tau^2}, \beta_{\tau^2}) \\ 1/r &\sim Unif(\alpha_r, \beta_r) \end{aligned} \quad (7)$$

ここで  $\mu_\beta, \Sigma_\beta, \alpha_{\sigma^2}, \beta_{\sigma^2}, \alpha_{\tau^2}, \beta_{\tau^2}, \alpha_r, \beta_r$  はハイパーパラメータで、実際には  $\mu_\beta = \mathbf{0}, \Sigma_\beta = 1000\mathbf{I}, \alpha_{\sigma^2} = 2, \beta_{\sigma^2} = 2, \alpha_{\tau^2} = 2, \beta_{\tau^2} = 0.1, \alpha_r = 3, \beta_r = 30$  と設定した。

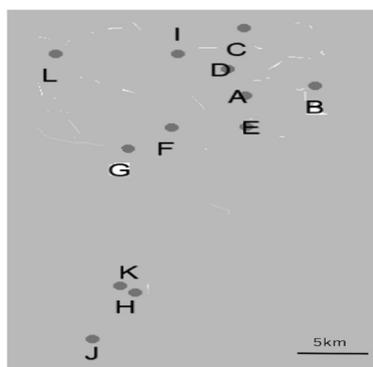


図2 各店舗の所在地。提供されたデータの店舗は同一地方に存在する。

### 3.3.2 モデルの推定方法

モデルの推定はマルコフ連鎖モンテカルロ法(Markov Chain Monte Carlo; 以下 MCMC と略記する)で行う。推定にはマルコフ連鎖にて 15,000 回のサンプリングを行い、はじめの 10,000 回をバーンイン期間として破棄し、残りの 5,000 個の MCMC 標本に基づいて得た事後統計量の平均値を推定値とした。

## 4. 実証分析

### 4.1 データについて

本研究では、経営科学系研究会連合協議会主催、平成 29 年度データ解析コンペティションで提供されたデータを使用する。内容は、2015 年 7 月 1 日から 2017 年 6 月 30 日までのヘアサロン 12 店舗における施術記録とショップでの販売商品の購買記録である。図 1 に各店舗の売上(店舗規模が特定されないように縦軸は非表示)を示し、図 2 にその所在地を示す。図 1 から店舗によって売上が大きく異なる様子が見て取れる。

式(1)の売上モデルの構築に際して、対象期間で入手可能なデータを用いて説明変数を整備する必要性から、主に都道府県、市町村および鉄道各社が公開するデータを利用している。本研究では「店舗周辺地域の人口[万人]」、「店舗周辺地域の男女比[%]」、「最寄り駅の乗車人数[万人/日]」、「最寄り駅までの距離[分]」、「出店する店舗のフロア[階]」を利用している(表1)。交通条件を表現するために、「最寄り駅までの距離」、「出店する店舗のフロア」の属性を用いている。店舗付近の商圈の大きさを表すために「店舗周辺地域の人口」、「店舗周辺地域の男女比」、「最寄り駅の乗車人数」を使用した。ここで最寄り駅の乗車人数は最寄り駅の各路線の合計値とし、店舗が存在する地域の人口は市町村レベルのデータを用いた。そのため同じ地域に複数店

表 1 推定に利用する説明変数

記号	説明変数	取得方法
$y_{s_j}$	店舗の年間総売上	提供データ
$x_{s_j,1}$	店舗周辺地域の人口 [万人]	オープンデータ [19, 20]
$x_{s_j,2}$	店舗周辺地域の男女比	オープンデータ [19, 20]
$x_{s_j,3}$	最寄り駅の乗車人数 [万人/日]	JR・私鉄各線が公表する乗車人数 [20-22, 24]
$x_{s_j,4}$	最寄り駅までの距離 [分]	店舗 HP のアクセスに記載された値
$x_{s_j,5}$	出店する店舗のフロア [階]	店舗 HP のアクセスに記載された値

表 2 推定結果

説明変数	2.5%	50.0%	97.5%
切片	-2.42	8.25	19.05
店舗周辺地域の人口	-0.01	0.01	0.03
店舗周辺地域の男女比	-13.82	-4.36	5.67
最寄り駅の乗車人数	0.75	2.08	3.26
駅から店舗までの距離	-0.12	0.01	0.14
出店する店舗のフロア	-0.05	0.07	0.20
$\sigma^2$	0.28	0.70	2.73
$\tau^2$	0.01	0.04	0.17
$r$	3.15	6.85	28.66

表 3 通常クリギングとの比較

手法	RMSE	exp(RMSE)
通常クリギング	0.427	1.53
本研究の手法	0.053	1.05

舗が存在する場合は同一の値となる。

## 5. 推定結果とその活用

### 5.1 パラメータの推定結果

はじめに店舗 A から店舗 H の 2016 年度の POS データを利用したパラメータの推定結果を表 2 に示す。本論文では 2016 年度のみデータを利用した結果を示すが、2015 年度も同様に推定することが可能である。表中には事後平均と標準偏差の目安として 2.5% と 97.5% の領域を示している。推定結果から「最寄り駅の乗車人数」が 95% 信頼区間の意味で有意である。これは外部アンケート結果 [25] の「現在利用しているサロンの継続理由 Top10」に男女とも「交通の便がよい」が含まれていることから妥当な結果であると判断する。ただし共分散関数のパラメータの推定値から誤差の 94.5% ( $=\tau^2/(\sigma^2+\tau^2)$ ) は空間相関成分であることから推定に利用した説明変数では説明できていない要因が多く存在する（たとえば自宅から美容室までの距離、価格帯など）。今回は取得可能なオープンデータを利用したが、新規店舗計画の際の地域調査データ（競合店の有無、道路の通行量など）を用いることで改善を図ることが可能である。ただし、この点は今後の課題とする。

### 5.2 店舗の売上予測と従来手法との比較

式 (1) による売上高予測の内挿精度を 12 分割交差検定 (12-fold cross-validation) で検証した。

通常の 12 分割交差検定法は、データ全体を無造作に

12 組に分割し、「11 組のデータを用いてパラメータ推定を行い、残り 1 組のデータを予測して精度検証を行う」という過程を 12 回繰り返し予測精度を評価する手法である。本研究では POS データを各店舗の年間売上高に変換したデータに対して適用した。

精度検証にあたり比較対象手法として通常クリギングである

$$y(s) = w(s) + \epsilon$$

を利用した。通常クリギングの推定も 3.3.2 節と同様に MCMC にて 15,000 回のサンプリングを行い、はじめの 10,000 回をバーンイン期間として破棄し、残りの 5,000 個の MCMC 標本に基づいて得た事後統計量の平均値を推定値とした。

12 分割交差検定による売上高の予測精度を表 3 に示す。表中に示す値は  $\ln$  (予測した売上高) と  $\ln$  (POS から算出した売上高) の平均二乗平方根誤差 (RMSE) を計算し、 $\exp(\text{RMSE})$  と変換した値を示す。 $\exp(\text{RMSE}) = 1.0$  ならば予測した売上高と POS から算出した売上高が完全に一致することを表し、 $\exp(\text{RMSE}) = 1.1$  ならば予測した売上高と POS から算出した売上高が平均的に 10% 異なる状態を示す。

表 3 より、本手法が通常クリギングより高い精度で予測できる可能性をもつことがわかる。推定誤差も 5% であることからモデルの予測性能が高く実務での適用が期待される。ただし前節のパラメータ推定において有意である変数が一つであること、誤差分散が大きいことから過剰適合している可能性がある。過剰適合を検証するには他都道府県のデータなどを用いてさらなる検証が必要であるが、本論文ではデータが提供されていない都合上、他地点の売上を予測するが検証は今後の課題とする。

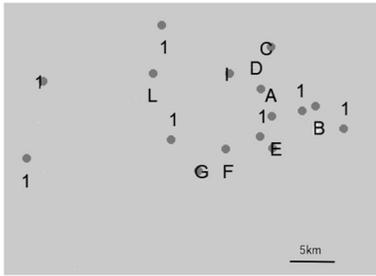


図3 モデルの予測結果  
色が濃いほど高い値を示す。図中のAからLはPOSデータが存在し、そこから1と記した地点の店舗の売上を予測した。

データ提供企業はほかにも店舗を有する企業であるため、参考までに首都圏のデータ未提供の店舗の売上高の予測結果を図3に示す。予測結果を示すと事業規模が明らかになるため、地図中にマッピングした。正解データが存在しないため結果の良し悪しは不明であるが、新規店舗出店の売上高を統計モデルと2種類のデータ（POSデータ、オープンデータ）から予測が可能となり、新規出店店舗の候補地を選出する手助けとなる。

## 6. おわりに

本研究では、複数店舗を経営している企業の新規出店店舗の売上予測を行う方法論として空間相関を構造化して内挿を実行できる普遍型クリギングに着目し、12店舗のヘアサロンのPOSデータに対して実証実験を行い、普遍クリギングの売上予測（内挿）の適用可能性を検証した。

実証実験の結果、首都圏近郊において予測誤差が5%程度である高精度な売上予測が可能であることが明らかになり、任意の地点の売上予測を提供できる可能性があることを確認した。また売上に寄与する要因の影響度をモデルの大域変動項である説明変数の回帰係数を推定することで明らかにした。

本研究では地球統計学や経済学などの他分野で利用されているクリギングを応用し、マーケティング分野においてその有用性を検証したことで、他分野の手法が利用できることを示し応用可能性を広げることができた。

今後の課題と今後の方向性として以下が挙げられる。本研究で用いたデータは複数店舗を経営する美容室の購買履歴データであるが、推定する対象のエリアに比べて、パラメータを推定するための店舗数は少ないといえる。そのためより新規店舗を計画しているエリア

近郊に既存店舗の購買履歴データが存在しない場合は予測精度が低下してしまう。そのような条件下において精度の高い予測を行うには本研究で利用したオープンデータ以外に新規店舗計画の市場調査データ（競合店舗数、近隣の開発状況など）をモデルに組み込むことが考えられる。式(5)に示すとおり空間的な相関を、付録に示すセミバリオグラムモデル化によって定式化している都合上、空間的な相関が正となっている。これはマーケティングにおいて強い制約であり店舗間の競合状況を説明できない（付録参照）。そのため式(4)の分散共分散に店舗間の競合等をモデル化することが考えられる。また表2の推定結果からわかるとおり、本モデル化では売上に寄与する要因が十分に抽出されていないことから改良が必要である。

さらに、店舗には複数年にわたって蓄積された購買履歴データが存在することから、距離の関数として定義していた共分散を時間と距離の関数、時空間共分散関数で定義することが挙げられる。たとえばCressie [12]では時間・空間で独立に共分散関数を定義し、それらの和や積の合成から時空間共分散関数を定義できることが示されている。時空間を含めたクリギングの推定のためのソフトウェアは数多く存在しており、RのパッケージではGraler et al. [26]にまとめられている。このように共分散関数の時間的な変動を取り入れたモデル化は今後の課題である。またアンケート結果からわかるように店舗への来店には交通の便が重要である。これらを取り入れるために共分散関数に異方性を取り入れることも今後の課題として挙げられる。たとえばBeauchamp et al. [16]は鉄道網の構造をモデルに取り込むことで等方的なモデルと比べてよりクリアな予測を可能とし、予測精度を高めている。

謝辞 本研究に利用したデータは経営科学系研究会連合協議会より提供いただきました。運営に関わる皆様とデータを提供いただきました会社に御礼申し上げます。また匿名のレフェリー2名からは大変有用かつ建設的なコメントいただきました。この場をお借りして御礼を申し上げます。

## 参考文献

- [1] 厚生労働省, 「平成28年度衛生行政報告例の概況」, <http://www.mhlw.go.jp/toukei/saikin/hw/eisei-houkoku/16/> (2018年6月25日閲覧)
- [2] 矢野経済研究所, 「理美容市場に関する調査を実施(2017年)」, <https://www.yano.co.jp/press/press.php/001684> (2018年6月25日閲覧)
- [3] V. Mahajan, E. Muller and F. M. Bass, “New prod-

uct diffusion models in marketing: A review and directions for research,” *Diffusion of Technologies and Social Behavior*, N. Nakicenovic and A. Grübler (eds.), Springer, pp. 125–177, 1991.

[4] V. Mahajan and E. Muller, “Innovation diffusion and new product growth models in marketing,” *Journal of Marketing*, **43**, pp. 55–68, 1979.

[5] F. M. Bass, “A new product growth for model consumer durables,” *Management Science*, **15**, pp. 215–227, 1969.

[6] 山口類, 土屋映子, 樋口知之, “状態空間モデルを用いた飲食店売上上の要因分解,” *オペレーションズ・リサーチ: 経営の科学*, **49**(5), pp. 316–324, 2004.

[7] W. J. Reilly, *Law of Retail Gravitation*, Wiley, 1931.

[8] D. R. Bell, H. Teck-Hua and C. S. Tang, “Determining where to shop: Fixed and variable costs of shopping,” *Journal of Marketing Research*, (**35**), pp. 352–369, 1998.

[9] D. Huff, “Defining and estimating a trading area,” *Journal of Marketing*, (**28**), pp. 34–38, 1964.

[10] D. Huff, *Determination of Intra-urban Retail Trade Areas*, Retail Estate Research Program, 1962.

[11] S. Yang and G. M. Allenby, “Modeling interdependent consumer preferences,” *Journal of Marketing Research*, **40**, pp. 282–294, 2003.

[12] N. Cressie, *Statistics for Spatial Data*, Wiley, 1993.

[13] S. Banerjee, B. P. Carlin and A. E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, 2004.

[14] 間瀬茂, 武田純, 『空間データモデリング』, 共立出版, 2001.

[15] 井上亮, 清水英範, 吉田雄太郎, 李勇鶴, “時空間クリギングによる東京 23 区・全用途地域を対象とした公示地価の分布と変遷の視覚化,” *GIS-理論と応用*, **17**, pp. 13–24, 2009.

[16] M. C. Beauchamp, C. de Fouquet and L. Malherbe, “Dealing with non-stationarity through explanatory variables in Kriging-based air quality maps,” *Spatial Statistics*, **22**, pp. 18–46, 2017.

[17] J. Li and A. D. Heap, “A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors,” *Ecological Informatics*, **6**, pp. 228–241, 2011.

[18] J. Valente, S. S. Wu, A. Gelfand and C. F. Sirmans, “Apartment rent prediction using spatial modeling,” *Journal of Real Estate Research*, **27**, pp. 105–136, 2005.

[19] 東京都総務局統計部, 「東京都の統計」, <http://www.toukei.metro.tokyo.jp/> (2018 年 6 月 25 日閲覧)

[20] 横浜市政策局統計情報課, 「横浜市統計ポータルサイト」, <http://www.city.yokohama.lg.jp/ex/stat/> (2018 年 6 月 25 日閲覧)

[21] JR 東日本, 「各駅の乗車人員」, <http://www.jreast.co.jp/passenger/> (2018 年 6 月 25 日閲覧)

[22] 小田原急行電鉄, 「1 日平均駅別乗降人員」, <https://www.odakyu.jp/company/railroad/users/> (2018 年 6 月 25 日閲覧)

[23] 東急電鉄, 「乗降人員」, <http://www.tokyu.co.jp/company/business/railway/passengers/index.html> (2018 年 6 月 25 日閲覧)

[24] 東京メトロ, 「各駅の乗降人員ランキング」, [https://www.tokyometro.jp/corporate/enterprise/passenger\\_rail/transportation/passengers/index.html](https://www.tokyometro.jp/corporate/enterprise/passenger_rail/transportation/passengers/index.html) (2018 年 6 月 25 日閲覧)

[25] 株式会社リクルートライフスタイル, 「20–69 歳男女の美容サロン利用実態」, [https://hba.beauty.hotpepper.jp/wp/wp-content/uploads/2015/06/cencus\\_fullreport\\_201506.pdf](https://hba.beauty.hotpepper.jp/wp/wp-content/uploads/2015/06/cencus_fullreport_201506.pdf) (2018 年 6 月 25 日閲覧)

[26] B. Graler, E. Pedesma and G. Heuvelink, “Spatio-Temporal Interpolation using gstat,” *The R Journal*, **8**, pp. 204–218, 2016.

[27] D. Allard, J.-P. Chilès and P. Delfiner, *Geostatistics Modeling Spatial Uncertainty*, 2nd edition, Wiley, 2013.

## 付録

### 分散共分散関数とセミバリオグラム

空間過程における分散共分散関数の構造化のために誤差項が従う分布が弱定常性をもつと仮定する。弱定常性は任意の  $\forall \mathbf{s}, \mathbf{h}$  に対して次式が成立すると仮定される。

$$E[Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})] = \mathbf{0},$$

$$\text{Var}[Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})] = 2r(\mathbf{h}).$$

ここで  $2r(\mathbf{h})$  はバリオグラム,  $r(\mathbf{h})$  はセミバリオグラムと呼ばれる関数である。バリオグラムはその性質として,

$$r(\mathbf{0}) = 0,$$

$$r(\mathbf{0}) \geq 0,$$

$$r(-\mathbf{h}) = r(\mathbf{h}),$$

を満たし, 2 次定常性をもつ確率場は本質定常性をもつといえる。

任意の地点における確率変数の空間予測を行う際, 予測誤差の分散は定数  $a_1, \dots, a_N$  を用いて

$$\text{Var} \left[ \sum_{i=1}^N a_i Y(s_i) \right] = \sum_{i=1}^N \sum_{j=1}^N a_i a_j C(s_i - s_j) \quad (8)$$

として与えられ, 式 (8) はゼロ以上を満たす必要がある。すなわち分散共分散関数の要素は正であるという条件を満たすように構成される必要がある [27]。このように地球統計モデルにおいては分散共分散関数は対応するセミバリオグラムによって表現できる。そのため分散共分散関数をモデル化するよりもセミバリオグラムを用いてモデル化するほうが推定精度に関する点でよい [12]。