特集　不動産統計情報の充実をどう図るか（現状と課題）

不動産統計情報と計算機の利用
―データ利用のための課題―

株式会社おたに　　小谷　祐一朗
おたに　ゆういちろう
全日本不動産協会　　内田　健太郎
うちだ　けんたろう

概要

「オープンデータ」と呼ばれるデータの公開形式により、様々な不動産に関する統計情報が政府や地方自治体から公開されるようになった。このような情報により、不動産に関するデータを高度化させ、不動産の実務にも役立てることができる。また、「Web スクレイピング」と「オープンデータ」について、データの収集方法としての可能性についても述べる。さらに、不動産統計情報のデータの前提について述べ、計算機を使った計算方法である通常のブートストラップ法と「巨大なデータ」を扱うためのブートストラップ法の一つである Bag of Little Bootstrap 法の比較も行う。

キーワード：ラビーネット、計算統計、ブートストラップ法、Bag of Little Bootstrap（BLB）、モンテカルロ法、リサンプリング、シミュレーション、空間分析、オープンデータ、確率過程

はじめに

近年、政府や地方自治体から様々な統計データが公開されており、それには不動産に関連するデータも多い。公開されたデータを利用して、不動産に関する情報を高度化することが期待されている。物件情報が多様化し、不動産に関する経済指標が新たに作成されることで、不動産に対して様々な見方ができるようになる。これにより、不動産事業者だけでなく、消費者も安全・安心な取引を期待できるようになるであろう。

当然だが、データの利用を行うためにはデータを収集しなければならない。インターネットの発達に伴い、不動産に関するデータを効率的に集めることが可能になり、経済指標に組み込むこともできる。ただし、いくつかの考慮すべき点もあるため、これらを把握して適切に利用することが重要である。

また、データの実体はあくまで数字や文字、記号の集まりであるから、データには何らかの文脈や意味づけが伴う。つまり、データは何らかの事象及びそれが生ずる世界（空間）を具体的に表現するものである。また、データの分析には計算機の高度利用は不可欠である。従って、不動産統計情報の利用には、データの生成する背景を考慮し、さらには計算機の特徴を踏まえた分析方法の利用や開発が必要である。

以上の観点から、本稿では、まず、不動産統計情報の概要について述べる。次に、不動産統計情報を収集する方法について説明する。最後に不動産統計情報の理論的背景と計算機を使った計算方法の比較結果も紹介する。

不動産統計情報とその実務利用

「不動産統計情報」とは不動産に関する統計データであるが、その範囲は広い。具体的にはどのようなものがあり、どのような種類に分けること

表１：不動産に関するデータの分類例

	集計・加工なし	集計・加工あり
直接的	地価公示（国土交通省） 都道府県地価調査（都道府県） 路線価（国税庁）	不動産価格指数（国土交通省） 不動研住宅価格指数（不動産研究所）
間接的	国勢調査（総務省） 位置参照情報（国土交通省）	消費者物価指数（総務省）

ができるのであろうか。本項では、不動産統計情報の分類を行い、その形式と量の観点で整理を試みる。さらに、不動産の流通実務において取り扱われているデータ及びデータの利用における課題についても述べる。

まず、不動産統計情報には「ローデータ」や「個票」といった集計・加工が行われていないデータと、データそのものを加工・集計して作成される指数等のデータがある（表 1）。これらのデータには、建物や土地といった不動産を直接表現するものと人口や住所等の不動産を間接的に表現するものがある[1]。前者は国土交通省の「地価公示」や都道府県の「都道府県地価調査」、国税庁の「路線価」等があり、後者は総務省の人口に関する調査である「国勢調査」や国土交通省の住所とその位置に関する「位置参照情報」等がある。また、作成される指数等の代表例には、不動産を直接的に表現するものとしては国土交通省の「不動産価格指数[2]」や不動産研究所の「不動研住宅価格指数[3]（旧東証住宅価格指数）」があり、間接的に表現するものには総務省統計局の「消費者物価指数」等の経済指標がある。

これらは「属性」「時間」「空間」あるいはその組み合わせで、統計データとして特徴付けることができる。例えば、単年度の地価公示は地価公示点の属性を持つクロスセクションデータである。また、不動産価格指数等は時間を軸とした時系列データである。地価公示や都道府県地価調査等は緯度経度等の情報を持つ空間データであり、複数年度の地価公示はクロスセクションと時系列の双方の特徴を持つパネルデータである。更に、時系列データと空間データの双方の特徴を持つ時空間データと捉えることもできる。

次に、このようなデータを量の視点でみてみよう。公開されている不動産に関する統計データから単純なデータセットを作成する場合（例えば、地価公示と国勢調査を組み合わせる場合）、最終的なデータの量（行や列の数）は決して多くはならない。一方、得られるデータに基づいて地理空間に関する分析を行う場合[4]、データの数（すなわち行数や列数）が１列増えるだけで、計算量が急激に増加することがある。つまり、単純に量が少ないデータでも、潜在的なデータに目を向けると、非常に巨大なデータとなる可能性が不動産統計情報にはある。

では、このようなデータは不動産の実務ではどのように扱われており、どのような課題があるのだろうか。ここでは、全日本不動産協会の「ラビーネット不動産（https://rabbynet.zennichi.or.jp/）」を例に考えてみよう。

まず、全日本不動産協会では、公正な不動産取引の実現、不動産取引における消費者の安全の確

[1] ただし、不動産統計情報の定義域やその分類の境界は非常に曖昧である。

[2] 国土交通省「不動産価格指数」http://www.mlit.go.jp/totikensangyo/totikensangyo_tk5_000085.html

[3] 川口有一郎、渡部光章（2011）「取引価格データベースを用いた住宅価格指数」http://www.reinet.or.jp/pdf/fudoukenjutakuhyouka/data05-20151215.pdf

[4] 例えば、クリギング等が挙げられる。クリギング実行時に用意するデータセットは、データとそのデータの取得地点（つまり座標）だが、その計算過程では地点間の地理的距離を全て算出する。

図１：ラビーネット不動産における物件情報の提供例

保や不動産の有効利用の促進等を目的として、不動産業界の健全な発展に寄与する活動及び社会貢献活動を行なっている。社会貢献活動には「災害の被災者等の支援活動」「地域社会の健全な発展に資する啓発活動」等がある。このような背景から、「ラビーネット不動産」は適正な物件情報の流通及び取引の推進を目的としており、扱われるデータは会員事業者が提供する個別の物件に関するデータである。

一方で、ラビーネット不動産で提供する情報は、データの形式や量の観点から、拡張の余地は十分にある。特に、国土交通省や総務省等の政府機関や地方自治体が提供するデータと会員の事業者が保有する情報を有機的に結合することで、事業者と消費者の双方にとって有益な情報を創出することもできるであろう。この場合は、提供する情報の質や継続可能性についての検討が必要になる。また、情報の種類によっては、迅速かつ正確に消費者に届ける方法も検討しなければならない。

そのためには、政府や地方自治体が、提供するデータの種類や量を充実させるだけでなく、会員事業者のデータに関する知識の習得やデータの取扱技術(データリテラシー)の研鑽も重要になる。例えば、データの収集方法や利用方法及びそれに伴う問題点等についても具体的に学んでいく必要がある。

従って、公開されている不動産統計情報には様々な種類・形式のものがあり、これらを有機的に活用していくことが重要だと考えられる。また、不動産の実務においては、現時点ではいわゆるローデータを扱うことが多い。ただし、データリテラシーを課題として、それを向上させることで、各種のデータを利用した様々な指標等を開発・提供していくことも考えられるであろう。

データの収集方法

データの利活用を行うためには、データを収集し、どのような利用方法があるかを検討しなければならない。データの収集には様々な方法があるが、近年では「Web スクレイピング（スクレイピング)」と「オープンデータの利用」が比較的容易な収集方法として知られている。ここで、スクレ

イピングとは「プログラムによって、取得したいデータが公開されているページに HTTP（Hyper Text Transfer Protocol）アクセスを行い、必要な情報をページから抽出し、自動でデータ収集を行う[5]」ことであり、「オープンデータ」とは「著作権や特許等の制御メカニズムなしに自由に利用できる」形式のデータのことである。本項では、この二つの方法の利用とその問題点を紹介する。

まず、スクレイピングによって収集した不動産に関するデータを経済指標の作成に利用する試みがある。例えば、ベルギーの統計局では、2020年にWebスクレイピングしたデータを消費者物価指数（CPI）へ組み込むための研究が行われており、組み込む項目の一つに「rent for student rooms」がある。スクレイピングによって収集する理由として以下の4つ（要約）が示されている[6]。

1. 以前はCPIの対象だったが、集計に非常に時間がかかるため除外されたこと
2. 取引管理のデータベースに登録がなかったり、他の取引との区別が難しいこと
3. プライバシーや学生特有の契約期間（10-12ヶ月）によりデータの収集が難しいこと
4. 世帯の予算において極めて高くなることが推計されたこと

つまり、スクレイピングによって収集するデータは組み込み対象である経済指標（CPI[7]）に影響を与える可能性があるが、他の調査方法では収集が難しい。そのため、スクレイピングを利用していることがわかる。スクレイピングは、このような収集が困難であったり、旧来の方法では時間がかかるデータを集めるためには非常に効率が高く有用な方法である。

ただし、スクレイピングには問題点もある。まず、収集したデータはバイアスの存在を前提としなければならないこともある。例えば、集められるデータはあくまでWeb上のものであり、現実世界のものとは異なることがある[8]。また、データの収集そのものに継続性リスクが存在する。スクレイピングの対象となるWebサイトの運営企業が何らかの方法でプログラムによるアクセスを拒否した場合、その継続的な利用（例えば経済指標の作成）は限定的にならざるをえない。さらに、過度なアクセス等はネットワークの帯域とサーバーに負荷がかかる。イギリスの国家統計局の資料[9]によれば、過度な負荷をかけるようなアクセスはDDoS攻撃[10]のようなものであり、風評被害（reputational risk）も存在するとされている。また、法的問題として、著作権やデータベースに関する権利等も挙げられている。

他方、不動産統計情報の収集に「オープンデータ」を利用する場合がある。特に政府等の官公庁が提供するオープンデータは、各種の指標のローデータとして利用されていることもある。例えば、国土交通省は地価公示から「公示47住宅指数・商業指数[11]」を算出している。これは47都道府県の

---

[5] 似たような概念にクローリング（crawling）があるが、こちらはページ内リンクを辿り、様々なWebページにアクセスを行い、HTMLのソースコードをダウンロードするものである。従って、クローリングの場合は、不特定多数のWebページを対象とするが、スクレイピングの場合は特定多数のWebページが対象である。

[6] Loon, V. K. and Roels, D.（2018）“Integrating big data in the Belgian CPI”, *Meeting of the Group of Experts on Consumer Price Indices*, pp.19-20.
https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Belgium.pdf
訳は筆者によるものである。

[7] なお、通常の賃貸物件の賃料は組み込まれている。

[8] 加えて、ダイナミックプライシング等による価格付けが行われているWebサイトも存在する。

[9] Matt, G.（2017）“Better Scraping, Better Statistics? Using web-scraped data in statistical outputs”, Office for National Statistics.
https://gss.civilservice.gov.uk/wp-content/uploads/2018/01/Better-Scraping-Better-Statistics-1.pdf

[10] 大量のリクエストを送ることでWebサービスを利用不能にさせる悪意を持った攻撃のこと。もちろん、この攻撃はスクレイピングやクロールと目的は異なる。ただし、手法としては同じであるから、サービス提供者の視点からすればスクレイピングもDDoS攻撃も同じという考え方もある。

[11] 国土交通省（2018）「地価公示の都道府県（全国）最高価格地（住宅地・商業地）を用いた平均価格指数及びその変動率」
http://www.mlit.go.jp/common/001226814.pdf

住宅地・商業地の最高価格（継続地点）を利用した指数であり、オープンデータとして公開されている地価公示のデータのみで算出可能な指標の一つである。

また、その性質から、再利用を前提としたデータとして成型されていることも多い。ダウンロードだけでなく、API（Application Programming Interface）による形式で提供されている場合もある。例えば、総務省統計局が運営する「政府統計の総合窓口（e-Stat）」のAPI機能（https://www.e-stat.go.jp/api/）では、国勢調査や住宅・土地統計調査等の不動産統計情報が提供されている。データを提供することがAPI機能のそもそもの目的であり、ある程度の規模のアクセスも想定されている。

しかし、オープンデータにも問題は存在する。例えば、国勢調査等のデータでは、統計情報という観点からデータの粒度（行政区画等による地域差がなく、同一の細かさで分布する程度）が粗いこともあり、ビジネスなどで使用するための粒度を満たしづらいといったこともある。「誰でも利用可能」であり、かつ「個人を特定することができない」ためには、ローデータや個票のレベルで公開は難しい。そのため、データの粒度が問題として生じやすい。

ただし、これはデータ分析の方法等を工夫することで、限定的ではあるが、解決可能である。一例が株式会社おたにが運営するGEEO（https://geeo.otani.co/）である。各種オープンデータを利用して開発したGEEOでは、日本全国約6,000万地点の複数の不動産価格[12]を計算することができ、その一つに推定公示地価がある。

地価公示は約26,000地点しかなく、地価公示点が存在しない地域も存在する。従って、地価公示は全国で均一に分布している指標とは言い難い。

図2：GEEOの推定公示価格

一方、推定公示地価は、独自の方法で推定を行うことで日本中を同一の粒度で算出するための計算を行なっている。すなわち、確率論的補間等を踏まえた統計手法の開発により、粒度の不揃いの問題を推定値の正確性の問題に置き換えている。なお、GEEOで使用している統計モデルの自由度調整済R二乗は0.92である。

不動産統計情報を集めるためには「スクレイピング」や「オープンデータの利用」が考えられる。双方とも有用な方法だが、利用にあたっては検討すべき点もある。これらを踏まえて、データを収集することが、不動産統計情報の高度利用の最初の一歩であろう。

[12] 市場の取引に基づく推定価格（複合不動産）、地価公示に基づく推定公示地価（土地）、建築費などから算出する建物の価格と推定公示地価を足し合わせた推定積算価格があり、これらは推定時系列データとしても提供を行なっている。

データの前提

ところで、不動産に関するデータの収集やデータを利用した経済指標等の設計を行うためには、不動産に関する事象が生成される空間の存在が前提となる。それは、不動産の価格や不動産との距離、あるいはその関係性等が生成される抽象的な空間[13]であり、それは不動産に関するあらゆる事象を生成する元を持つ集合等と考えてもよい。すなわち「不動産統計情報全体の標本空間」であり、本項では$\Re$で表す。また、生成される複数（k個とする）の不動産統計情報の標本空間をΩとし、その部分集合をAとする。

つまり、

$$\bigcup_{k=1}^{\infty} \Omega_k \in \Re$$

$$\bigcup_{m=1}^{\infty} A_m \in \Omega_k$$

と考える[14]。

さらに、A_mの具体的なデータ（例えば地価公示）をXとした場合、

$$X \subseteq A_m$$

$$X = [x_1, x_2, \ldots, x_n]$$

である。なお、当然ながら、xは確率空間上の可測関数（確率変数）である。

このx、すなわち不動産統計情報の大きな特徴として、時間または空間といったパラメータにより変化することが挙げられる。

従って、これは確率過程であり、

$$X = [x(s_1), x(s_2), \ldots, x(s_n)]$$

と表せる。

ここでsは時間Tや空間Sあるいはその双方$S \times T$を意味する。当然のことながら、実際にデータを扱う際にはsは離散的$i = [1,2,3,\ldots]$であるが、本来は連続的な$i = [0, \infty)$である。また、Sはユークリッド空間R^nとして考えることが多いが、不動産が実際に存在する地球の曲率は0ではない。

また、実際に収集するデータには観測誤差ε（これもsによって特徴付けられる）が含まれるので、次のように書くことができる。

$$X = [x(s_1) + \varepsilon(s_1), x(s_2) + \varepsilon(s_2), \ldots, x(s_n) + \varepsilon(s_n)]$$

ただし、$\varepsilon(s_i) \sim i.i.d$

なお、$i.i.d$は独立同分布（independent and identically distributed）であり、平均0、標準偏差σの正規分布を仮定することが一般的である。

従って、不動産統計情報は空間や時間に従う確率過程として扱うべきものであるが、実際に扱えるデータは理論とは異なる部分が存在する。すなわち、不動産統計情報の高度利用には、データの前提を踏まえた分析や計算の方法を検討する必要があろう。

シミュレーションと計算資源

ところで、時間や空間等の確率過程を利用した分析を行う際には、モンテカルロ法等のシミュレーションによる計算を行うことがある。シミュレーションにおいて、計算量の少なさは、より多くの計算を行うためにも重要である。特に、不動産統計情報は潜在的なデータまで考慮した分析を行おうとすれば、計算量が多くなるため、従って、計算機の使用効率を高める必要がある。本項では平均値や中央値などの統計量をシミュレーションによって求める方法とその計算効率について説明する。

まず、単純に全国の無数にある地点の地価を母集団とし、地価公示$[x_1, x_2, \ldots, x_n]$をその標本と考えてみよう。標本平均は次の式で求める。

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}[x_1 + x_2 + \ldots + x_n]$$

nの数が大きいと、人間が手で計算するときも大変であるが、それは計算機も同じである[15]。平均

[13] 不動産が実際に存在する地理空間の意味ではない。なお、地理空間はこの空間が曲面として顕出しているものであり、地図はこの空間が平面として表現されているものと考えられる。

[14] これは数学記号を使って書き直しただけであるが、この空間を理論的に定義及び考察する道具として幾何学等が考えられる。

[15] 人間が数百から数千の要素を手で紙に書き出すのは大変である。計算機も数十億や数兆の要素を一度にメモリに読み出すのは大変である。

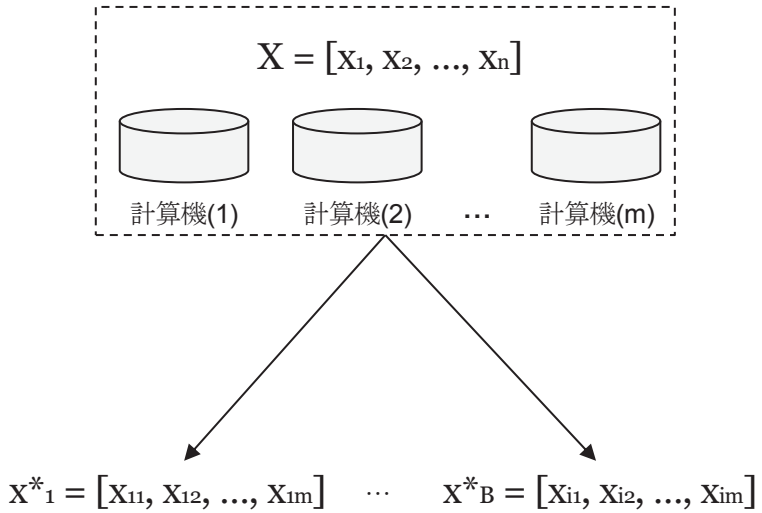

図 3：ブートストラップ標本と計算機の関係図

値を求める際に、ベクトルや行列の要素数が多い（つまりnの値が大きい）と、一台の計算機に搭載できるメモリの量では足りなくなり、別の処理を検討しなければならない。もちろん、これは標準偏差や回帰係数を求める場合等でも同じである。

この時の計算方法の一つに、古典的な手法ではあるが、「ブートストラップ法[16]」がある。これは、シンプルだが非常に強力な推定手法であり、「ブートストラップに伴う反復計算は並列化が容易」[17]な、再標本化による統計量の推定方法である。次の手順で行う。

1. 得られたn個のデータ$[x_1, x_2, \ldots, x_n]$をXとおく。
2. $[1,2,\ldots,n]$から等確率で整数を選ぶことをm回繰り返す。つまり、$[i_1, i_2, \ldots, i_m]$の整数列Kを生成する。なお、同じ整数を複数回選んでも良い（復元抽出）。
3. Kに基づき、Xから$[x_{i_1}, x_{i_2}, \ldots, x_{i_m}]$を取り出し、$x^*$とおく。
4. 2. と 3. をB回繰り返す。つまり$X^* = [x^*_1, x^*_2, \ldots, x^*_B]$である。
5. X^*のそれぞれから求めたい統計量を求める。

ここで、m台の計算機にデータセットである$X = [x_1, x_2, \ldots, x_n]$が分割されて配置されている場合（図3）を考えてみよう。

ランダムに各々の計算機からX^*の要素を繰り返して抽出し(つまりブートストラップ標本の生成)、そのそれぞれで求めたい統計量を求める。この処理は、別々の計算資源を割り当てることで並列化が容易となる。

巨大なデータに、このような方法を応用するためには計算資源を効率的に使うことが重要となる。Kleiner, etl.(2014)で提案された計算方法が「Bag of Little Bootstrap (BLB)」である。BLBは次の手順で行う。

Bag of Little Bootstrapの手順

1. 得られたn個のデータ$[x_1, x_2, \ldots, x_n]$をXとおく。
2. $[x_1, x_2, \ldots, x_n]$をm個のグループ$[x_{11}, x_{12}, \ldots, x_{1k}]$，$[x_{21}, x_{22}, \ldots, x_{2k}]$，…，$[x_{m1}, x_{m2}, \ldots, x_{mk}]$に分割する（非復元抽出）。
3. m個のグループ、それぞれでブートストラップ（復元抽出）を行い、m個の推定値$\Theta = [\theta_1^*, \theta_2^*, \ldots, \theta_m^*]$を得る。

[16] 「ブートストラップ」の名前は「Pull yourself up by your bootstrap.（自分で何とかせよ）」という英語のイディオムに由来する。ブートストラップ法は Efron（1979）によって提案された。

[17] 下平英寿（2011）「第8章 ブートストラップ」21世紀の統計科学 3 数理・計算の統計科学 第III部 統計計算の展開と統計科学 日本統計学会 HP版 2011年10月、pp.194（http://ebsa.ism.ac.jp/ebooks/sites/default/files/ebook/1881/pdf/vol3_ch8.pdf）

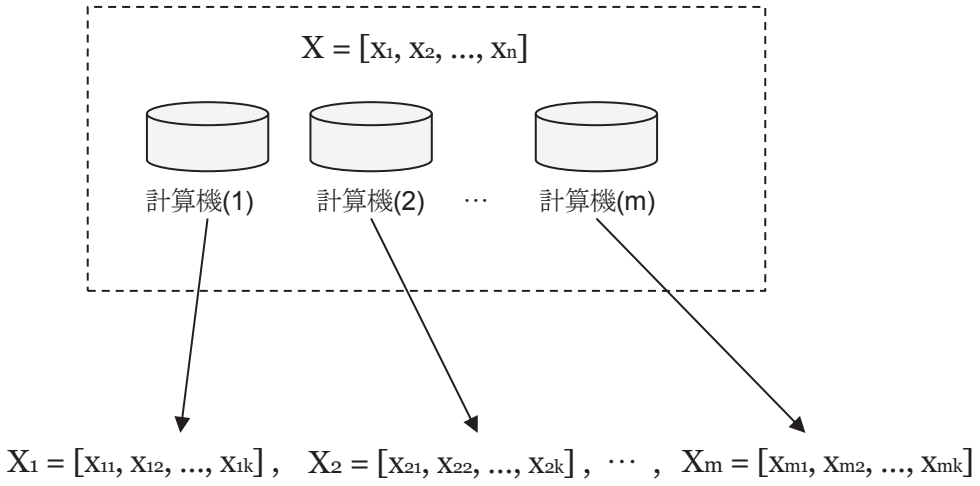

図 4：BLB 標本と計算機の関係図

4. Θの平均を計算する。

この場合、予めデータを取得する段階でデータを保持する計算機を定めれば 1 及び 2 はデータ取得のプロセスの一部となり、3 及び 4 のみが実質的な推定値を求めるための計算プロセスになる。図 4 は BLB を使用する場合の計算機と各サンプルの関係図である。

計算機毎に処理を分割できることが大きなメリットである。

図 5 は、非常に小規模なデータ（n = 25,988）ではあるが、自然対数をとった 2018 年の地価公示データに基づいて、ブートストラップ法によって発生させた 1,000 個のサンプルのヒストグラム及び BLB によって発生させた 1,000 個のサンプルのヒストグラムである（ただし、点線は実際のデータ（地価公示）の中央値）。ブートストラップ法の場合は 26 個のサンプルを発生させ中央値を求めることを 1,000 回繰り返した。また、BLB の場合は、1,000 個のサブセットに分割を行い、多項乱数に基づき各サブセットのサンプリング回数（ただし合計は 25,988）を決定した。双方とも 1,000 回ずつ標本抽出を繰り返し、そのそれぞれで平均値を求めた。なお、実際のデータの中央値は 11.205 である。

通常のブートストラップ法も BLB もあまり差がつかないことがわかる。ただし、BLB の場合はサブセット内の再標本に依存させる計算手法のため、計算資源を通常のブートストラップ法に比べて節約することができる。

Kleiner, etl.(2014)では「nが 1,000,000 のデータにおいて、… （中略 ）… 各データポイントが 1MB のストレージを使用している時、データの総量は 1TB になる。ブートストラップサンプリングの場合は約 632GB のスペースが必要になるが、BLB の場合多く見積もっても 4GB 程度である[18]」と述べている。このような場合における計算は通常は HDD(Hard Disk Drive)や SSD(Solid State Drive)等のストレージでなく、RAM(Random Access Memory)を使うことが容易となるため、計算時間が短縮されることも明らかである。

このように、計算機を利用したシミュレーションでは計算方法を工夫することで、計算資源の節約が可能になる。ただし、確率過程を扱うための手法としてはブロックブートストラップ法、計算精度を高めるための解析的な方法としてはサドル

[18] Kleiner, A., Talwalkar, A., Sarkar, P., & Jordan, M. I. A scalable bootstrap for massive data. Journal of the Royal Statistical Society, Series B (Statistical Methodology), Vol. 76 (2014), pp.795-816. 訳は筆者によるものである。なお、ここで「632GB」という数字の計算根拠は Efron and Tibshirai (1993) の 0.632n からである。

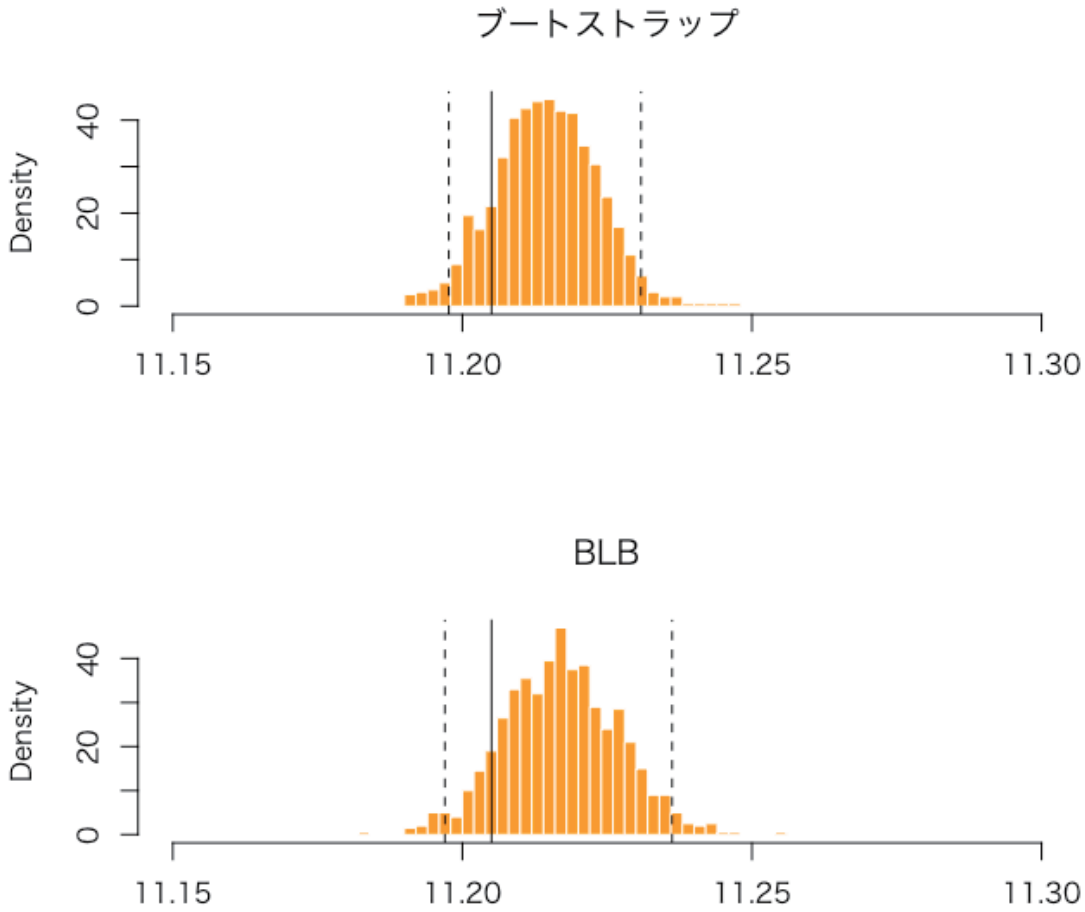

図 5：ブートストラップ標本（上）と BLB 標本（下）[19]

表 1：要約統計量

	最小値	2.5%点	中央値	平均値	97.5%点	最大値	標準偏差
ブートストラップ	11.191	11.198	11.214	11.214	11.236	11.246	0.009
BLB	11.184	11.197	11.217	11.217	11.231	11.255	0.010
地価公示（対数）	6.215	9.159	11.205	11.270	13.800	17.832	1.190

地価公示：N=25,988、ブートストラップ：N=1,000、BLB：N=1,000

なお、小数点第 4 位以下は四捨五入している。

[19] 実線は実際のデータ（地価公示）の中央値、点線はそれぞれの標本における 2.5%点と 97.5%点を表す。

ポイント近似等がある。従って、今後は不動産統計情報を分析するためにも、確率過程を踏まえた、計算資源を節約でき、かつ計算効率の良い手法等を開発することが必要であろう。

おわりに

さて、ここまで様々な論を展開してきた。不動産統計情報が充実し、様々な目的で利用できるようになることは、実務的な視点からも有用であると考えられる。特にオープンデータの形式で公開されているデータを利用した経済指標の作成と提供は事業者だけでなく、消費者にとっても大きなメリットがあるだろう。そのためにも、不動産統計情報を適切に収集し利用するためのデータリテラシーの向上や、計算機を利用した確率過程の計算手法のさらなる開発などが求められる。データの種類や量を充実させるだけでなく、データを扱うための技術や理論を発達させ普及させることが、不動産統計情報の利用の高度化を推進するためには肝要である。

参考文献

川口有一郎、渡部光章（2011）「取引価格データベースを用いた住宅価格指数」http://www.reinet.or.jp/pdf/fudoukenjutakuhyouka/data05-20151215.pdf

国土交通省（2018）「地価公示の都道府県（全国）最高価格地（住宅地・商業地）を用いた平均価格指数及びその変動率」 http://www.mlit.go.jp/common/001226814.pdf

下平英寿（2011）「第 8 章 ブートストラップ」21 世紀の統計科学 3 数理・計算の統計科学 第 III 部 統計計算の展開と統計科学 日本統計学会HP版 2011年10月（http://ebsa.ism.ac.jp/ebooks/sites/default/files/ebook/1881/pdf/vol3_ch8.pdf）

瀬谷創，堤盛人（2014）『空間統計学』，朝倉書店.

盛川仁，丸山敬（2001）『条件付確率場の理論と応用』，京都大学学術出版会.

Bakker, Bart F. M., Rooijen, Van J. & Toor, Van L.（2014）The System of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics. Statistical Journal of the IAOS, 30, pp. 411-424.

Bertail, P. and Politis, N. D.（2001）Extrapolation of subsampling distribution estimators: the i.i.d. strong mixing cases. Canadian Journal of Statistics Vol. 29, No. 4, pp. 1-14

Bickel, P. J., Freedman, D. A.（1981）Some Asymptotic Theory for the Bootstrap. Annals of Statistics. 9, No. 6, 1196-1217.

Callegaro M., Yang Y.（2018）The Role of Surveys in the Era of “Big Data”, The Palgrave Handbook of Survey Research. https://link.springer.com/content/pdf/10.1007%2F978-3-319-54395-6_23.pdf

Daniels, H. E.（1954）Saddlepoint Approximations in Statistics. Annals of Mathematical Statistics. 25, No. 4, pp. 631-650. https://projecteuclid.org/euclid.aoms/1177728652

Efron, B.（1979）Bootstrap Methods: Another Look at the Jackknife. Annals of Statistics Vol. 7, No. 1, pp. 1-26.

Efron, B. and Tibshirani, R.（1993）An Introduction to the Bootstrap. Chapman and Hall.

Horowitz, J. L.（2003）Bootstrap Methods for Markov Process. Econometrica. Vol. 71, No. 4, pp. 1049-1082. https://www.ssc.wisc.edu/~bhansen/718/Horowitz%20Markov%20Bootstrap.pdf

Kleiner, A., Talwalkar, A., Sarkar, P., & Jordan, M. I.（2014）A scalable bootstrap for massive data. Journal of the Royal Statistical Society, Series B (Statistical Methodology), Vol. 76, pp. 795–816.

Lahiri, S. N. and Zhu, Jun.（2006）Resampling Methods for Spatial Regression Models under a Class of Stochastic Designs. Annals of Statistics Vol. 34, No. 4, pp. 1774-1813.

Loon, V. K. and Roels, D.（2018）“Integrating big data in the Belgian CPI.” Meeting of the Group of Experts on Consumer Price Indices. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Belgium.pdf

Matt, G.（2017）Better Scraping, Better Statistics? Using web-scraped data in statistical outputs. Office for National Statistics. https://gss.civilservice.gov.uk/wp-content/uploads/2018/01/Better-Scraping-Better-Statistics-1.pdf

Rosen, S.（1974）Hedonic prices and implicit markets: product differentiation in pure competition. The Journal of Political Economy. 82, pp. 34-55.