モバイル空間統計　人口推計　統計処理

Technology Reports

社会・産業の発展を支える「モバイル空間統計」
―モバイルネットワークの統計情報に基づく人口推計技術とその活用―

モバイル空間統計における人口推計技術

先進技術研究所　寺田 雅之（てらだ まさゆき）　永田 智大（ながた ともひろ）
小林 基成（こばやし もとなり）

携帯電話ネットワークの運用データを用いて作成される人口統計であるモバイル空間統計について，推計の考え方と推計方式の概要を解説する．運用データからの人口の推計にあたっては，携帯電話ネットワークの特性に起因する推計値の偏りに留意する必要がある．また，基地局エリアごとに推計された人口をそのまま出力しても活用は困難である．そこで，モバイル空間統計の集計処理では，携帯電話ネットワークの特性を考慮した推計を行うことにより推計人口の偏りを除去するとともに，地理集計区分の変換プロセスを備えることによりメッシュなどの応用分野で利用しやすい形式で人口を出力することを可能とする．

1. まえがき

モバイル空間統計は，携帯電話ネットワークの運用データから継続的に生成される，日本全国の人口の統計情報である．

モバイル空間統計は，①非識別化処理，②集計処理，③秘匿処理，の3段階のプロセスを経て作成される．本稿では，非識別化された運用データから（メッシュごと・行政区画ごとなどの）集計エリアごとの人口を推計する，モバイル空間統計の中核部分である「集計処理」について，その推計の考え方と推計方式の概要を解説する．

2. 課　題

運用データからのモバイル空間統計の作成は，ごく単純化すれば以下の手順により行われる．

①運用データに基づいて，それぞれの基地局エリアに存在する携帯電話の台数を集計する．

②人口当りのドコモの携帯電話の普及率に基づいて，携帯電話の台数を人口に拡大する．

③上記で得られた基地局エリアごとの人口を，メッシュや市区町村ごとの人口へと再集計する．

しかし，携帯電話ネットワークは，そもそも人口統計を生成することを目的として設計されていないため，実際には上記のような単純な手順で運用データからモバイル空間統計を作成することはできない．

例えば，携帯電話の基地局は，着信制御のためにその基地局の通信圏内（基地局エリア）に在圏する携帯電話の存在をほぼ定期的（約1時間ごと）に確認している．運用データの一部である携帯電話の在圏データは，具体的にはこの在圏確認の状況を集約したデータである．しかし，在圏確認のタイミングは携帯電話ごとにばらばらであり，また，ある時点で基地局エリアに在圏して

© 2012 NTT DOCOMO, INC.
本誌掲載記事の無断転載を禁じます．

いることが確認された携帯電話が，次の確認時点までずっとその基地局エリアに在圏しているとは限らない．つまり，携帯電話ユーザの移動に伴って他の基地局エリアに移っているかもしれないが，ある確認時点から次の確認時点まで，その携帯電話が実際にはどの基地局エリアに在圏していたかについては，携帯電話ネットワークは知ることはできない．さらに，実際の携帯電話ネットワークでは，この確認の間隔は必ずしも一定ではなく，携帯電話の移動や利用の状況などによって大きく変動することがある．したがって，「ある時刻に何台の携帯電話が基地局エリアに在圏していたか」という情報を携帯電話ネットワークは直接知ることはできず，それぞれの基地局における在圏確認の状況から統計的に推定する必要がある．

また，基地局エリアごとの携帯電話の台数（在圏数）が推定できたとしても，これは当然ながらドコモの携帯電話ユーザ以外も含めた人口とは異なる値である．一見すると，日本の人口1人当りのドコモ携帯電話の普及率がわかれば，在圏数を普及率で単純に「割り戻す」ことにより人口を推計できるように考えられる．例えば，日本の人口が1億2000万人，ドコモの携帯電話サービスの契約数が6000万契約であるとする．このとき，人口2人に対して1台の割合（普及率＝0.5）でドコモの携帯電話が普及していることになるので，基地局エリアの在圏数を0.5で割り戻す（2倍する）ことにより，その基地局エリアに存在する人口を算出できる．しかし，実際には地域や年代・性別ごとにドコモの携帯電話の普及率は異なるため，このような単純な割り戻し処理では，地域や年代・性別によって大きく偏りが出てしまい，人口統計としては不適切なものとなる．さらに，携帯電話の電源は，常にオンの状態にあるとは限らない．例えば，電源が切れている携帯電話がどの基地局エリアに存在するかについて，携帯電話ネットワークが知ることはできない．そこで，これらの要因を考慮したうえで，携帯電話の在圏数を人口へと適切に拡大推計する必要がある．

最後に，これらの課題を解決して基地局エリアごとの人口を正しく推計できたとしても，それだけでは人口統計として活用することは困難である．例えば，「基地局Xのエリアの人口はY人である」という単位で人口を提供されても，そもそも基地局Xというのがどこにあって，どのくらいの広がりをもつエリアなのかがわからなければ，この人口データに使い道はない．さらに，携帯電話の基地局には屋内局／屋外局，利用する無線周波数帯などさまざまなバリエーションがあり，それぞれの基地局でエリアカバー範囲も大きく異なる．そのため，基地局エリア単位で推定された人口を，都市計画や防災計画など，人口統計の活用分野で良く使われる単位である，メッシュや行政区画（市区町村界など）単位の人口へと変換する過程が必要となる．

以上をまとめると，集計処理は以下の3段階の推計プロセスから構成される．

①在圏数推計処理

非識別化された運用データに基づき，運用データのばらつきなどを考慮しつつ，基地局エリアごとの携帯電話の在圏数を推計する．

②拡大推計処理

在圏数推計処理の結果である基地局エリアごとの携帯電話の在圏数に基づき，ドコモ携帯電話の普及率の偏りや携帯電話の電源断の影響などを考慮しつつ，基地局エリアごとの人口へと拡大推計する．

③エリア変換処理

拡大推計処理の結果である基地局エリアごとの人口に基づき，基地局エリア単位に推計された人口をメッシュや行政区画単位など，応用分野で活用しやすい集計単位へと変換する．

3. 集計処理

在圏数推計処理，拡大推計処理，エリア変換処理のそれぞれについて，基本的な推計の考え方と推計方式を示す．

3.1 在圏数推計処理

在圏数推計処理の目的は，非識別化された運用データに基づき，ある時間帯にそれぞれの基地局エリアに何台の携帯電話が在圏していたか，その在圏数を推計することである．

前述のとおり，携帯電話ネットワークは「ある時点で，どの基地局に何台の携帯電話が在圏しているか」を正確に把握しているわけではない．そこで，1時間ごとなどのある一定期間（以下，観測期間Tと呼ぶ）に収集された在圏データ[*1]に基づき，その期間の平均在圏数を求める．観測期間Tにおける，基地局エリアCの平均在圏数をm_C^Tとする．在圏数推計は，それぞれの基地局エリアごとにm_C^Tを推計する．

在圏数m_C^Tは，観測期間Tにおける，基地局エリアCでの携帯電話の滞在時間の総和を，観測期間の長さで除したものとして計算できる（図1）．すなわち，観測期間T，基地局エリアCにおける携帯電話iの滞在時間をt_{iC}^Tとすれば，

$$m_C^T = \Sigma_i t_{iC}^T \ / \ |T| \qquad (1)$$

として在圏数を求めることができる．なお，$|T|$は観測期間Tの長さを表す．

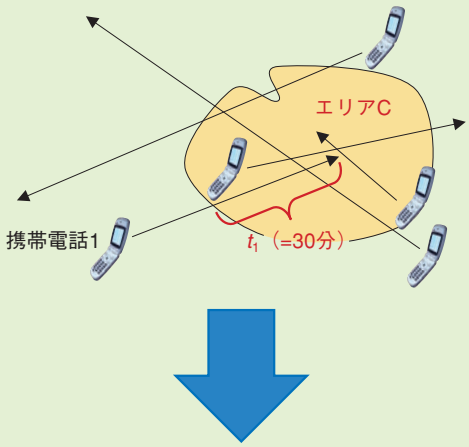
しかし，携帯電話ネットワークは前述のとおり常に携帯電話の在圏状況を把握しているわけではないため，それぞれの携帯電話の正確なエリア滞在時間t_{iC}^Tはわからない．そこで，在圏データの生成状況から統計的にこれを推定する．

もし，在圏データの生成契機である在圏確認の間隔がある周期ωで一定であれば，エリア滞在時間がω増えるごとに，そのエリアで観測される在圏データの数が1つ増えることが確率的に期待される．したがって，基地局エリアCにおいて，期間Tに観測された在圏データの集合をL_C^Tとすると，エリア滞在時間の総和$\Sigma_i t_{iC}^T$の期待値は式(2)のように推定できる．

$$E(\Sigma_i t_{iC}^T) = \omega |L_C^T| \qquad (2)$$

しかし，2章で述べたように，実際の携帯電話ネットワークでは，在圏確認の間隔は一定でなく，携帯電話の利用や移動の状況などのさまざ

・たとえば，ある1時間に，「基地局エリアC」を5台の携帯電話が通過もしくは滞在したとする．
・ここで，携帯電話1～5は，この1時間のうちに，それぞれ t_1～t_5 の時間だけエリアCに在圏していたとする．
・たとえば，携帯電話1の在圏時間は，t_1=30分だったとする．



在圏数推計の考え方：
携帯電話 i は，1時間（60分）の間にそれぞれ t_i（分）在圏していた．
⇒上記1時間のある時点において，携帯電話 i がエリアCに在圏している確率は，t_i/60である．
（たとえば，携帯電話1が在圏している確率は 30/60 = 1/2）
⇒すなわち，エリアCに在圏する携帯電話数（在圏数）の期待値は，
t_1/60+ t_2/60+… t_5/60 = Σ t_i/60（台）となる．

図1　在圏数推計の考え方

＊1　**収集された在圏データ**：正確には，非識別化処理により，在圏データと属性データの組合せに対し，氏名や電話番号などの個人を識別する情報を除去するとともに，住所や生年月日などの属性を居住地コードや年齢層などへと丸めることにより非識別化された運用データである．処理のイメージをわかりやすくするために，以降では在圏データと呼称する．

まな要因により動的に変化する．すなわち，個々の在圏データごとにωが変動する（在圏データの時間密度が変化する）．そこで，これらの在圏データの時間密度のばらつきを考慮し，それぞれの在圏データに対して周辺の信号間隔に応じた特徴量（重み）を与える．ある在圏データに対する特徴量は，典型的には前後の在圏データとの間隔の平均として与えることができる．ある在圏データjの特徴量をω_jとすると，エリア滞在時間の総和$\Sigma_i t_{iC}^{T}$は式(3)で推定できる．

$$\mathrm{E}(\Sigma_i t_{iC}^{T}) = \Sigma_{j \in L_C^T} \omega_j \qquad (3)$$

したがって，観測期間Tにおける，基地局エリアCの在圏数m_C^Tは，当該期間およびエリアで観測された在圏データL_C^Tとその特徴量により，式(4)を用いて推計される．

$$\mathrm{E}(m_C^T) = \Sigma_{j \in L_C^T} \omega_j \,/\, |\mathrm{T}| \qquad (4)$$

ここまでの議論では，人口構成や移動人口の推計に必要となる，属性（年齢層・性別・居住エリア）ごとの在圏数推計には言及してこなかった．ただし，これらの属性ごとの在圏数も，上記の属性なしの式に対する簡単な変更のみで推計することができる．

ある（年齢層，性別，居住エリア）の組からなる属性条件をAとする．例えば，東京都在住の30代男性を表す属性は，A＝（30代，男性，東京都）と表される．このとき，前述のt_{iC}^{T}のうち，属性Aの携帯電話iの滞在時間を$t_{iC,A}^{T}$とすれば（$t_{iC,A}^{T} \subseteq t_{iC}^{T}$），

$$m_{C,A}^T = \Sigma_i t_{iC,A}^{T} \,/\, |\mathrm{T}| \qquad (5)$$

として，東京都在住30代男性の，上記エリア・時間帯における在圏数を求めることができる．したがって，$L_{C,A}^T$を在圏データL_C^Tのうち属性Aの携帯電話に関する在圏データとすれば，これの期待値は式(6)で推計できる．

$$\mathrm{E}(m_{C,A}^T) = \Sigma_{j \in L_{C,A}^T} \omega_j \,/\, |\mathrm{T}| \qquad (6)$$

3.2 拡大推計処理

拡大推計処理では，在圏数推計処理で求めた在圏数（携帯電話数）に基づいて人口を推計する．

前述のとおり，もしドコモの携帯電話の人口当り普及率が，地域や年齢層・性別などによらず一様であり，かつすべての契約ずみ携帯電話の電源が入っておりどこかの基地局エリアに必ず在圏しているという前提を置くことができれば，拡大推計処理では単に推計された在圏数に対して，普及率の逆数を乗ずる（割り戻す）のみで人口を推計することができる．

しかし，現実には，年齢や性別，居住地域によってドコモの携帯電話の普及率は大きく異なる．この普及率の偏りを正しく反映することなしには，推計された人口も大きく偏りを生じることになる（例えば，実際の普及率が高い地域では人口が過大に，普及率が低い地域では過少にそれぞれ偏って推計される）．

さらに，人々が就寝する夜間帯などで携帯電話の電源が切られているとすれば，その間は携帯電話と携帯電話ネットワークの間のやりとりは行われないため，これらの携帯電話はどの基地局エリアにも在圏数として表れない．したがって，単に契約状況だけに基づいて拡大の乗数（拡大率と呼ぶ）を決定すると，実際の人口に対して，特に在圏率が小さい（電源OFFの端末が多いなど）時間帯において，実際よりも過少に人口が推計されることになる．

そこで，拡大推計では性別・年齢，居住地などの属性によって変わる普及率を正しく人口推計値に反映させるために，属性ごとに拡大推計処理を行う．また，拡大率の算出にあたっては，単に普及率を用いるのではなく，携帯電話の在圏状況を加味した在圏率を用いる．**図2**に拡大推計処理のイメージ図を示す．

具体的には，まず，属性ごとの普及率の偏りを正しく反映させるために，国勢調査人口や住民基本台帳人口などの公的統計による在住人口を属性ごとに集計する．この属性Aの在住人口をr_Aとする．属性の分割単位は，本稿執筆時点では，年齢5歳区分，性別ごと，都道府県ごととしている．次に，電源OFFなどによる非在圏端末の影響をなくすため，観測期間Tにおいて，それぞれの属性に相当する総在圏数を（全国の基地局エリアから集約して）求める．観測期間Tにおける全国の属性Aの総在圏数$m_{*,A}^T$の期待値は，式(7)により推計できる．

$$\mathrm{E}(m_{*,A}^T) = \Sigma_C \mathrm{E}(m_{C,A}^T) \qquad (7)$$

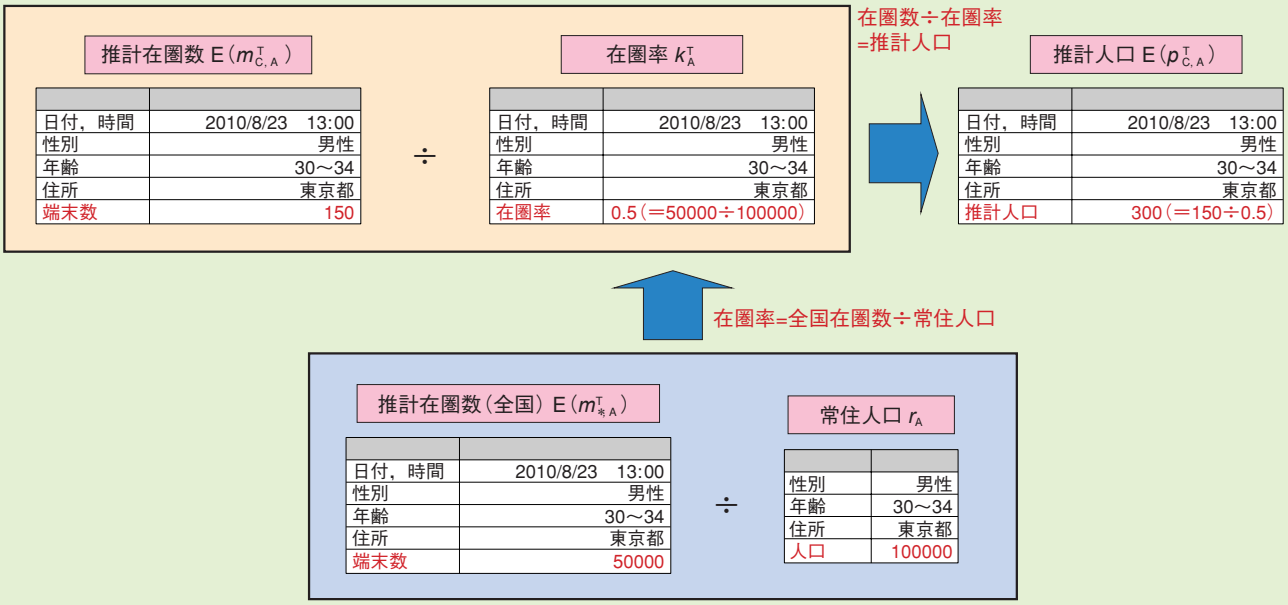


図2　拡大推計処理のイメージ図

このとき，観測期間Tにおける属性Aの人口について，人口1人当りの在圏携帯電話数，すなわち在圏率k^{T}_{A}は，式(8)で求められる．

$$k^{T}_{A}=E(m^{T}_{*,A})/r_{A} \tag{8}$$

このとき，観測期間T，基地局エリアCにおける属性Aの人口$p^{T}_{C,A}$の期待値は，式(9)によって推計される．

$$E(p^{T}_{C,A})=E(m^{T}_{C,A})/k^{T}_{A} \tag{9}$$

上記観測期間，基地局エリアにおける人口p^{T}_{C}は，当該期間・エリアの属性ごと人口の総和として求められる．

$$E(p^{T}_{C})=\Sigma_{A}E(p^{T}_{C,A}) \tag{10}$$

このように，拡大推計処理では，性別・年齢・地域でのドコモ携帯電話の普及率の偏りや，携帯電話の電源状態などによる在圏数の変動による影響を軽減し，基地局エリアごとの携帯電話在圏数から人口を算出する．

3.3 エリア変換処理

エリア変換処理は，拡大推計処理までによって推計された基地局エリアごとの人口を，500m/1kmなどの各サイズのメッシュごと，もしくは市区町村などの行政界ごとなどの各種の応用分析で一般に用いられる集計単位ごとの人口に変換する．**図3**にエリア変換処理のイメージ図を示す．

エリア変換処理では，基地局エリアごとに集計された人口を，実際の基地局エリアの大きさや形状などに応じてメッシュなどの集計単位に配分する．すなわち，観測期間Tにおける集計単位Gの人口p^{T}_{G}の期待値は，以下のように推計される．

$$E(p^{T}_{G}) = \Sigma_{C}d_{C\to G}E(p^{T}_{C}) \tag{11}$$

ここで，$d_{C\to G}$は基地局エリアCから集計単位Gへの人口の配分比であり，基地局エリアごとにその総和は1となる．すなわち，$\Sigma_{C}d_{C\to G}=1$が成立する．

配分比$d_{C\to G}$は，具体的には基地局エリアと各集計単位との地理的な「重なり」の大きさに基づいて求められる（面積按分と呼ばれる）．基地局エリアCの面積を$|C|$，基地局エリアCと集計単位Gとが地理的に重なる面積を$|C\cap G|$とする．このとき，配分比$d_{C\to G}$は以下の式により求められる．

$$d_{C\to G}=\frac{|C\cap G|}{|C|} \tag{12}$$

これにより，基地局エリアごとに集計された人口は，集計エリアと基地局エリアの重なり具合に応じて，

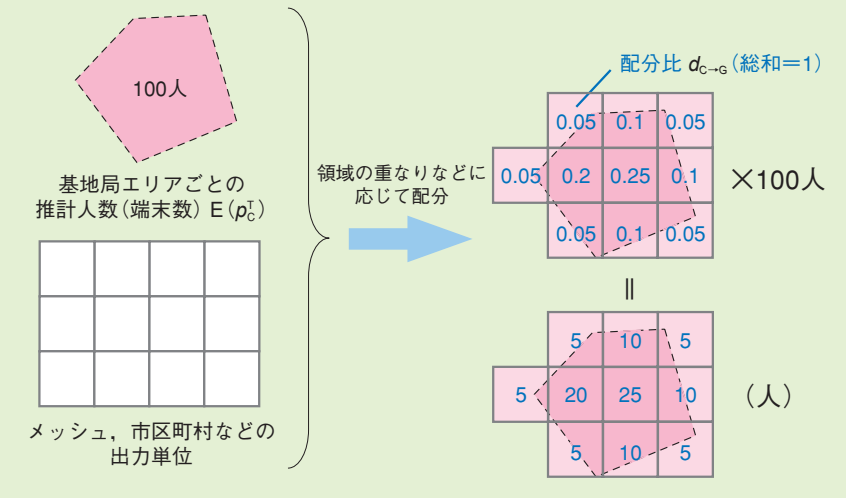


図3　エリア変換処理のイメージ図

各種の応用分析でより使いやすいメッシュなどを単位とした人口に変換される．

4. あとがき

本稿では，モバイル空間統計において人口を推計するための中核部分となる集計処理について，その課題を示すとともに，課題を解決するための基本的な推計の考え方と推計方式を解説した．

ただし，本稿で説明した内容は，あくまで集計処理における基本方式であり，実際にはさらにモバイル空間統計の信頼性を向上させるための各種の補正処理を経てモバイル空間統計は作成されている．これらの補正処理の例としては，携帯電話ネットワークにおける規制や故障が，人口の推計値に与える影響を軽減するための補正や，エリア変換処理の精度を向上させるための，基地局エリアの推定範囲の補正などが挙げられる．

これらの補正処理については，本稿執筆時点でも検証・改善の途上にあるものもあり，応用分野におけるニーズを見据えながら，より信頼性が高く，応用分野におけるより重要な意思決定を支えるのに足るモバイル空間統計を提供できるよう研究開発を進めていく．