小地域推定の概要

東京大学
空間情報科学研究センター
時空間社会経済システム部門
丸山 祐造
maruyama@csis.u-tokyo.ac.jp

CSISシンポジウム
2001.09.26

小地域推定とは

全国規模の多くの標本調査の目的
標本の背後にある母集団の持つ
情報の的確な縮約

調査区
市区町村などの細かい地域(小地域)

全国の母集団だけでなく，小地域の母集団(部分母集団)の特質の把握も標本調査の重要なテーマ!

何が問題か??

全体として標本数は十分であっても，
個々の地域毎には，標本の大きさは小

少標本から得られた通常の標本平均
推定誤差が大きい(後述)

推定量を安定させる(推定誤差を小さくする)ためには，どのような統計的手法を用いればよいか?

小地域，小標本 1

有限母集団からの非復元抽出
(超幾何分布のモデル)

大きさ N の有限母集団
各個体の特性値 $a_i,\ i = 1, \cdots, N$

$$\begin{array}{ccccc} 1 & 2 & 3 & \cdots & N \\ a_1 & a_2 & a_3 & \cdots & a_N \end{array}$$

例 首相を支持するかどうか?

$$a_i = \begin{cases} 1 & \text{支持} \\ 0 & \text{不支持} \end{cases}$$

小地域，小標本 2

推定したい値 (多くの場合，母平均)

$\mu = N^{-1}\sum_{i=1}^{N} a_i$ 例の場合: 首相の支持率

n 個の標本 $X_1, \cdots, X_n$ をもとに推定

自然な推定量 $\bar{X} = n^{-1}\sum_{j=1}^{n} X_j$

性質 $E(\bar{X}) = \mu$,

$Var(\bar{X}) = (N-n)(N-1)^{-1}\sigma^2/n$

分散は n の減少関数,

n が小さい程不安定

小地域，小標本 3

大きさ $2N$ の同質の母集団

標本数 $2n$，標本平均 $\bar{Y} = (2n)^{-1}\sum Y_j$

$Var(\bar{X})/Var(\bar{Y}) = (2N-1)/(N-1) > 1$

N が小さいほど推定量が不安定

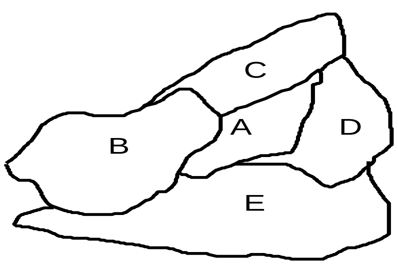
各地域で分散(推定の精度)を同等にするには，小地域からより多くの標本を得る必要

コストを考えると非現実的

推定量安定化のイメージ 1

・当該標本調査以外の情報の利用
(例 選挙での自民党の得票率)

・近接・隣接地域の標本の利用
↑ これを考える!



自然な発想

地域Aは，隣接地域BCDEと似てるはず

推定量安定化のイメージ 2

地域Aの母平均の推定

Aの標本平均をABCDE全体の標本平均に近づける.

修正率 n(A): 地域Aでの標本数

n(A)が小 大きく近づける

n(A)が大 あまり近づけない

このような推定量は，ベイズ統計学の枠組で導くことが出来る.

ベイズ統計学

- 母数 θ のとる値について，事前に情報が与えられていて，それが確率分布 $p(\theta)$ として表現 (事前分布)
- θ が与えられたもとでのデータ X の確率密度関数 $f(x|\theta)$
- $X=x$ が与えられたときの θ の条件付密度 (事後分布)
  $p(\theta|X=x) = (\int f(x|\theta)p(\theta)d\theta)^{-1} f(x|\theta)p(\theta)$
- ベイズ推定量 (事後分布の期待値)
  $\hat{\theta} = \int \theta f(x|\theta)p(\theta)d\theta / \int f(x|\theta)p(\theta)d\theta$

合計特殊出生率の例

平成5-9年度
人口動態保健所・市町村別統計(厚生省)

前書き
市町村別の統計に対して，出現数の少なさに起因する偶然性の影響のため，数値が不安定という問題
→ 地域間等の比較に耐え得るより安定性の高い指標をベイズモデルにより，算出

厚生省の報告では，統計的な解説が不十分なので，ここではそれを補う.

統計モデル

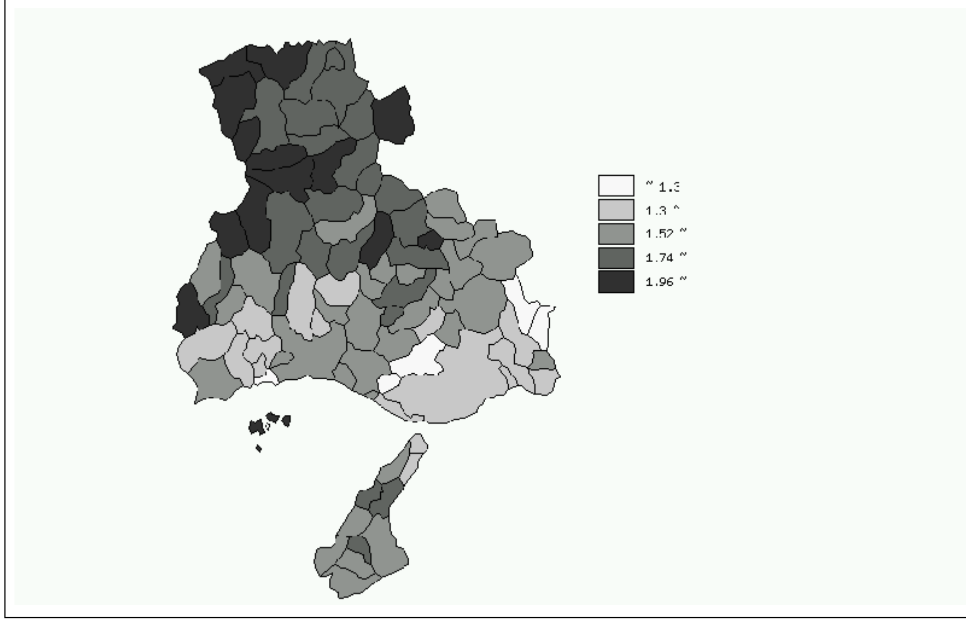
地域 i，年齢階級 j の出生数 X_{ij} が $(j=1,\cdots,7$: 15〜49歳を5歳毎に区切る$)$2項分布 $Bin(N_{ij},\theta_{ij})$ に従う.
母数 θ_{ij} は，出生率

i 地域の合計特殊出生率 $\sum_{j=1}^{7}\theta_{ij}$

最尤推定量 $\sum_{j=1}^{7} X_{ij}/N_{ij}$

N_{ij} が小さい場合，最尤推定量は不安定であることが知られている.

兵庫県の合計特殊出生率(最尤推定)



ベイズ推定量 1

事前分布 $\theta_{ij} \sim Be(\alpha_{Ij}, \beta_{Ij})$
厚生省の報告では，母集団が同質であると想定される「二次医療圏」なる区域に対し(例えば兵庫県では，91 市町村を 10 の二次医療圏に区分けしている)，共通の事前分布を与える．

θ_{ij} のベイズ推定量

$\hat{\theta}_{ij} = (\alpha_{Ij} + x_{ij})/(\alpha_{Ij} + \beta_{Ij} + N_{ij})$

従って合計特殊出生率のベイズ推定量

$\sum_{j=1}^{7} (\alpha_{Ij} + x_{ij})/(\alpha_{Ij} + \beta_{Ij} + N_{ij})$

ベイズ推定量 2

α_{Ij}，β_{Ij} の決め方

X の周辺分布 (尤度) を最大化する

$$\prod_{i=1}^{K} \binom{N_{ij}}{x_{ij}} \frac{\Gamma(\alpha_{Ij}+\beta_{Ij})\Gamma(\alpha_{Ij}+x_{ij})\Gamma(N_{ij}-x_{ij}+\beta_{Ij})}{\Gamma(\alpha_{Ij})\Gamma(\beta_{Ij})\Gamma(\alpha_{Ij}+\beta_{Ij}+N_{ij})}$$

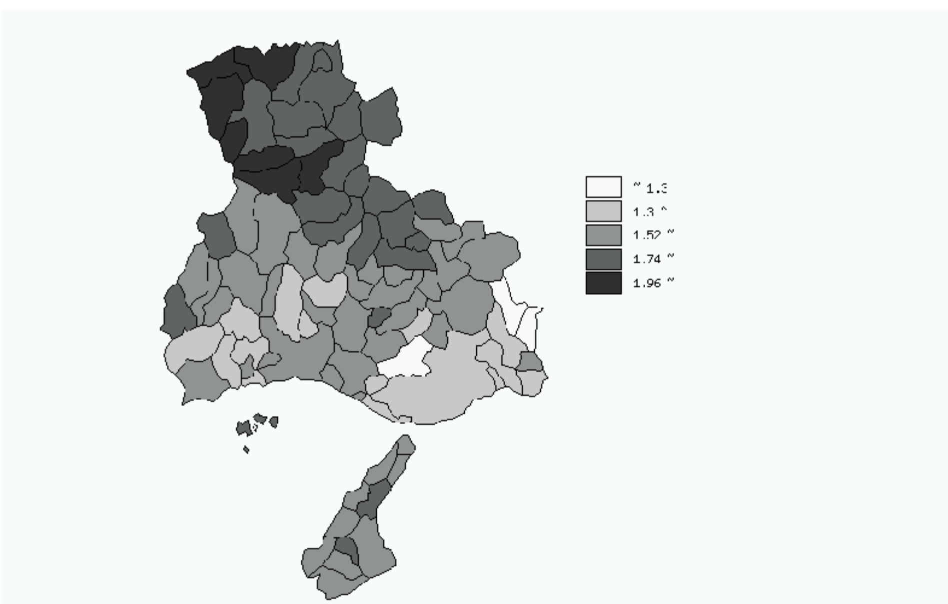
ベイズ推定量の解釈

$$\hat{\theta}_{ij} = \frac{\hat{\alpha}_j+\hat{\beta}_j}{\hat{\alpha}_j+\hat{\beta}_j+N_{ij}} \frac{\hat{\alpha}_j}{\hat{\alpha}_j+\hat{\beta}_j} + \frac{N_{ij}}{\hat{\alpha}_j+\hat{\beta}_j+N_{ij}} \frac{x_{ij}}{N_{ij}}$$

N_{ij} が小 $\rightarrow \hat{\alpha}_j/(\hat{\alpha}_j + \hat{\beta}_j)$ 二次医療圏全体での最尤推定値に近い．

N_{ij} が大 $\rightarrow x_{ij}/N_{ij}$

兵庫県の合計特殊出生率(ベイズ推定)



まとめ

小地域推定の重要性を数理統計的に解説

応用例: 厚生省の報告に用いられた経験ベイズ推定量の解説

小標本を部分母集団に読み替えることが出来る．