空間モデリング

古谷知之

講義内容

- 空間リスクの分析と可視化
- 確率地図
- ポアソン確率地図
- 相対リスク
- ポアソン・ガンマモデル
- 経験ベイズ推定値のMoran’s I

目的

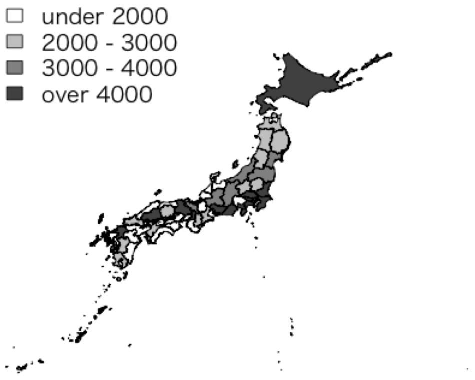
- イベント発生件数が非常に少なく、多くの地域（期間）で発生数が 0 となるようなデータを分析対象とする
  - 死因別死亡者数、事故発生件数など
- この場合、イベント発生件数を直接比較するのではなく、発生件数を人口数などで割った発生率で地域間比較をするのが望ましい
- 死亡や犯罪発生など、まれに発生するような空間現象を、確率値として比較したい場合には、粗率、相対危険度、ポアソン確率地図などが適用される

粗率

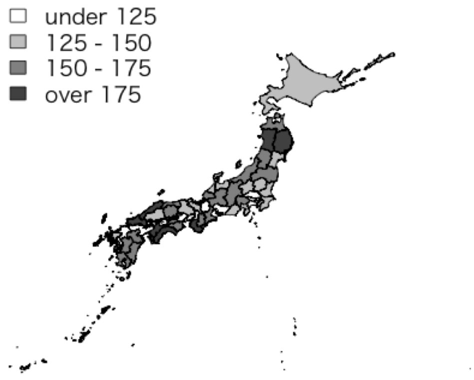
- 地区$i(=1,\ldots,N)$における、ある観測ケースの標本発生件数（例えば、心疾患死亡者数）をO_i、人口数をy_iとする
- このとき、人口数に対する観測ケースが発生する粗率r_iは以下のようになる

$$r_i = \frac{O_i}{y_i}$$

観測値と粗率



観測値



粗率

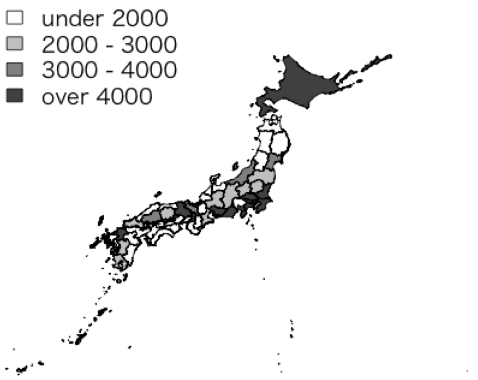
相対危険度

- 粗率r_iは対象地域全体の人口数や観測ケースの値を反映していないため、人口数に対する観測値の大小を判断することは容易でない
- そこで、対象地域全体で観測ケースの発生率が均一であると仮定して、観測ケースの標本値（観測値）O_iを期待値E_iで除した相対危険度θ_iを用いることがある
- 相対危険度は次式で定義される

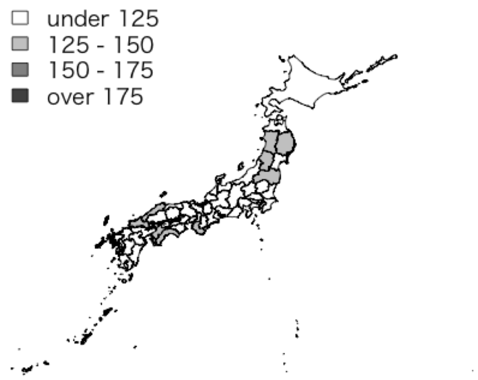
$$E_i = y_i \frac{\sum_{i=1}^{n} O_i}{\sum_{i=1}^{n} y_i}$$

$$\theta_i = \frac{O_i}{E_i}$$

期待値と相対危険度



期待値



相対危険度

ポアソン確率地図

- 観測ケースの発生頻度が低いものの、非常に多く観測回数が繰り返される場合、ポアソン確率地図を用いることで、人口規模による観測値の過大／過小評価の影響を考慮することができる
- 期待値E_iが与えられたとき、観測値O_iがポアソン分布に従うなら、以下のように表せる

$$O_i \sim Po(E_i)$$

- 期待値E_iを観測ケース発生の平均値μ_iとみなすと、地域iの観測数（推定値）がxとなるリスクp_iは次式で与えられる

$$p_i = \sum_{x \geq O_i} \frac{\mu_i^x exp(-\mu_i)}{x!} \; or p_i = \sum_{x < O_i} \frac{\mu_i^x exp(-\mu_i)}{x!}$$

ポアソン分布

- N個の点が互いに独立して任意の区画に同じ確率で現れる場合，点分布はポアソン分布となる

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

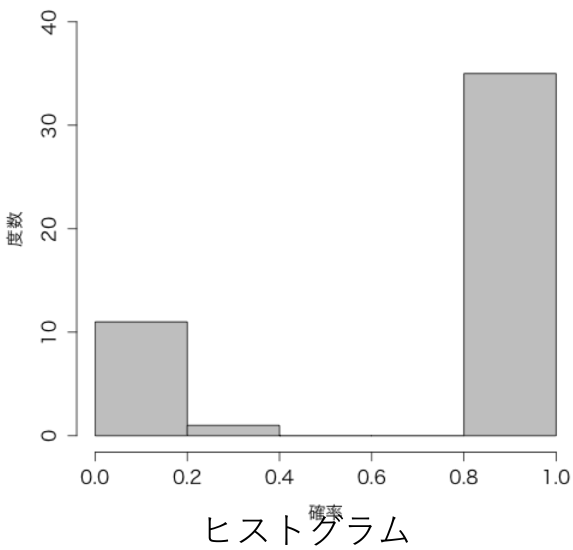
$$\lambda = \rho a = (N/A) \cdot (A/q) = N/q$$

a：区画面積＝対象地域の面積A/区画数q

ρ：単位面積あたりの平均点数（点密度）

ポアソン確率地図

ポアソン確率地図



ヒストグラム

相対リスクの経験ベイズ推定

- 観測数が少ない場合に、人口数の地域差を調整する方法の一つに、相対リスクの推定量$\hat{\theta}_i$を経験ベイズ推定する方法がある
- 相対リスクをベイズ推定する際には、相対リスクの分布を何らかの確率分布として仮定する
- その事前分布を規定する超パラメータを最尤推定法などにより推定する法と、マルコフ連鎖モンテカルロ法により事後分布を推定する方法とがある
- 超パラメータの分布を任意に（経験的に）与えベイズ推定する方法を経験ベイズ推定法といい、超パラメータ自体の分布を仮定して階層的に事後分布を推定する方法を、階層ベイズ推定法という

Marshallの方法による経験ベイズ推定

- 分析対象としている観測ケースの標本値O_iがポアソン分布に従って発生すると仮定できるようなケースであり、その期待値をE_iとする
- 相対リスクθ_iの最尤推定値をx_iとすると、θ_iが与えられた条件下での、x_i平均$E(x_i|\theta_i)$と分散$V(x_i|\theta_i)$はそれぞれ以下のように表せる

$$E(x_i|\theta_i) = \theta_i$$
$$V(x_i|\theta_i) = \theta_i/E_i$$

- 疫学分野では、 x_iを標準化死亡比（SMR）と呼ぶ

Marshallの方法による経験ベイズ推定

- θ_iの平均と不偏分散について事前情報を与えたとき、グローバルな経験ベイズ推定量$\hat{\theta}_i$はその事後情報として以下の収束推定量として求めることができる

$$\hat{\theta}_i = \hat{\mu} + \hat{C}_i(x_i - \hat{\mu})$$

- ただし

$$\hat{\mu} = \sum_{i=1}^{n} O_i \Big/ \sum_{i=1}^{n} E_i$$

$$\hat{C}_i = \frac{s^2 - \hat{\mu}/E_i}{s^2 - \hat{\mu}/\bar{E} + \hat{\mu}/E_i}$$

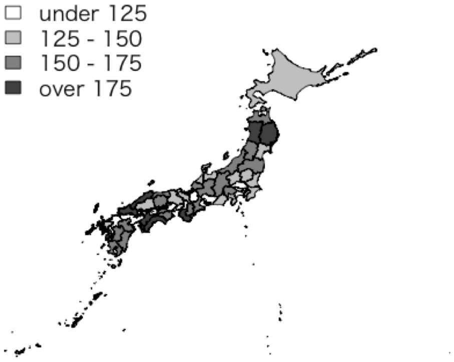
$$s^2 = \sum_{i=1}^{N} E_i(x_i - \hat{\mu}) \Big/ \sum_{i=1}^{N} E_i$$

Marshallの方法による経験ベイズ推定

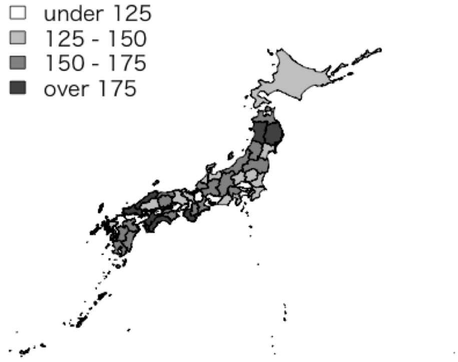
- 隣接行列を用いて、近隣地区との隣接性を考慮した経験ベイズ推定量を、ローカルな経験ベイズ推定量という
- 地区ijの隣接行列の要素 c_{ij}が与えられたとき、 $\hat{\mu}, \hat{C}_i, s^2, \bar{E}_i$を、それぞれ隣接要素を考慮して地区$i$毎に求める
- 例えば次式のようになる

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n} c_{ij} O_j}{\sum_{j=1}^{n} c_{ij} E_j}$$

Marshallの方法による経験ベイズ推定



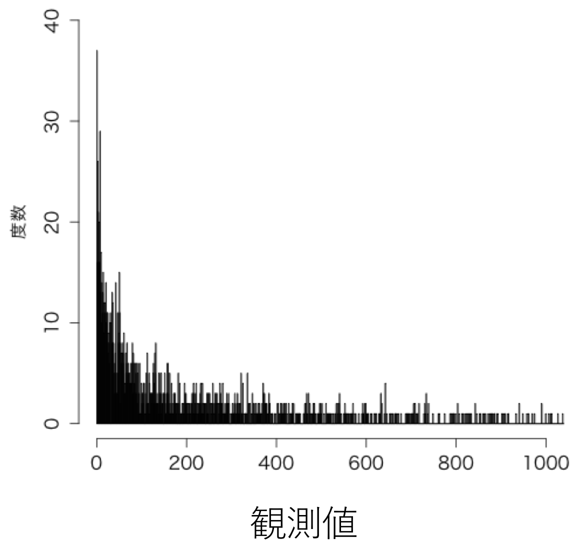
グローバルな経験ベイズ推定量



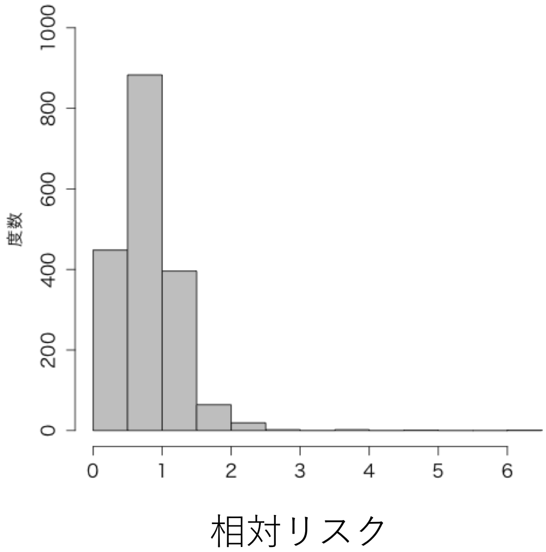
ローカルな経験ベイズ推定量

市町村別交通事故死者数の観測値と相対リスク

総務省統計局『社会生活統計指標-市区町村指標-』2007年の交通事故死者数データ



観測値



相対リスク

ポアソン・ガンマモデル

- 期待値E_iと相対リスクθ_iが与えられた条件付きで、観測値O_iがポアソン分布に従い、相対リスクがガンマ分布に従うとも考えられる
- このようなデータの相対リスクを推定する方法として、ポアソン・ガンマモデルが提案されている
- このモデルでは、相対リスクθ_iが観測値O_iと独立にガンマ分布に従って生成される

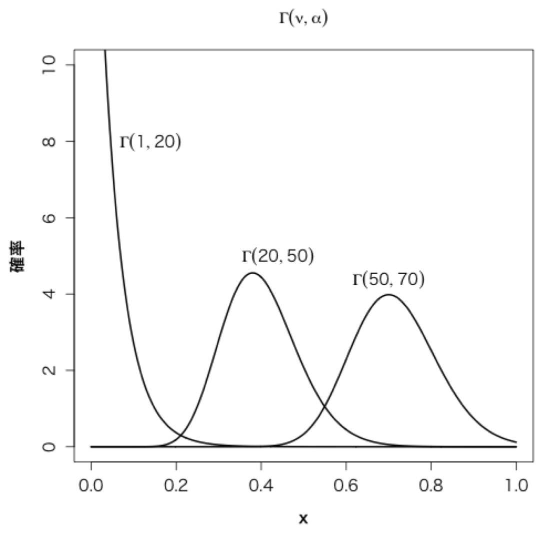
$$O_i \sim Po(\theta_i E_i)$$
$$\theta_i \sim \Gamma(\nu, \alpha)$$

ガンマ分布

- ガンマ分布の確率密度関数は、以下の式で表される

$$f(x) = \frac{1}{\Gamma(x)} \alpha^{\nu} x^{\nu-1} exp(-\alpha x)$$

- νは形状パラメータ、 αはスケールパラメータを意味する
- ガンマ分布の平均はν/α、分散はν/α^2となる



ポアソン・ガンマモデルの経験ベイズ推定

- 観測値O_iと期待値E_i及びガンマ関数のパラメータν, αを用いて、平滑化相対リスク$(O_i+\nu)/(E_i+\alpha)$を経験ベイズ推定できる
- パラメータν, αの事前情報を適当に与え、次の２式を繰り返し計算することにより、各パラメータの事後情報$\hat{\nu}, \hat{\alpha}$を計算し、相対リスクの事後分布を求める

$$\frac{\hat{\nu}}{\hat{\alpha}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i$$

$$\frac{\hat{\nu}}{\hat{\alpha}^2} = \frac{1}{N+1}\sum_{i=1}^{N}\left(1+\frac{\hat{\alpha}}{E_i}\right)\left(\hat{\theta}_i - \frac{\hat{\nu}}{\hat{\alpha}}\right)$$

ポアソン・ガンマモデルの経験ベイズ推定

- 相対リスクθ_iの最尤推定値x_iの事前分布がガンマ分布に従うとすると、その経験ベイズ推定量は、次式の収束推定量として求めることもできる

$$\hat{\theta}_i = \frac{E_i}{E_i+\alpha}x_i + \frac{\alpha}{E_i+\alpha}\frac{\nu}{\alpha}$$

対数正規モデル

- 相対リスクθ_iに多変量対数正規分布を考慮し、相対リスクの対数$log((O_i+1/2)/E_i)$をEMアルゴリズムやMCMC法によりベイズ推定する方法も提案されている
- 経験ベイズ法による対数正規相対リスクの事後分布b_iは次式のように表される

$$b_i = \frac{\hat{\phi} + (O_i + 1/2)\hat{\sigma}^2 log((O_i + 1/2)/E_i) - \hat{\sigma}^2/2}{1 + (O_i + 1/2)\hat{\sigma}^2}$$

$$\hat{\phi} = \frac{1}{N}\sum_{i=1}^{N} b_i$$

$$\hat{\sigma}^2 = \frac{1}{N}\left\{\hat{\sigma}^2 \sum_{i=1}^{N} [1 + \hat{\sigma}^2 (O_i + 1/2)]^{-1} + \sum_{i=1}^{N} \left(b_i - \hat{\phi}\right)^2\right\}$$