

# 自殺における相対リスクのベイズ推定

## —経験ベイズ法と階層ベイズ法による縮約推定量の構成—

紺田広明<sup>1</sup>

### 【要約】

効果的な自殺対策を行うために、地域における自殺実態の把握の重要性が認識されてきている。地域の自殺実態を表す統計指標として、相対リスクである標準化死亡比がしばしば用いられる。しかし、標準化死亡比は、人口が少ない地域において変動が大きい欠点がある。そのため、信頼性ある推定量が求められてきた。ここでは、最尤推定量、経験ベイズ推定量、階層ベイズ推定量の3種類の相対リスクの推定量を扱うこととした。兵庫県でのベイズ推定では、事前分布にガンマ分布を仮定した相対リスクの基本となる Poisson-Gamma モデルを用い、事前分布を仮定するベイズ推定の枠組みが縮約推定量を構成することを説明した。また、マルコフ連鎖モンテカルロ法による階層ベイズ推定量の構成の仕方を検討し、それぞれの推定量の特長についての整理を行った。

キーワード：自殺、標準化死亡比、階層ベイズ法、MCMC、小地域推定

## 1. はじめに

地域の自殺実態をあらわす指標として、標準化死亡比 (standardized mortality ratio : SMR) がしばしば用いられる。SMR は、端的には実際に観察された観測死亡数と期待死亡数との比である。期待死亡数は、基準 (全国など) の死亡率を、対象地域に当てはめた場合に期待される死亡数である。このことにより、この SMR は、相対リスク (relative risk) と呼ばれることがある。

通常、相対リスクは、リスク比 (risk ratio) や発生率比 (incidence rate ratio) などの「暴露群のリスクは、非暴露群のリスクの何倍であるか」という相対的な効果を表す用語として使用される (Gordis, 2009)。全国平均を基準とする SMR は、観測された死亡数と期待される死亡数の比と定義されるので、この値が 1 以上の場合は全国平均よりも高く、1 以下の場合は低いと相対的に判断することができる場所に特徴がある<sup>2</sup>。

SMR が自殺に関する統計処理に使用される主な理由は、自殺リスクへの年齢の交絡を統制している指標であるからである。一般に死亡のリスクに最も影響を与える要因は年齢であり、自殺のリスクも年齢と強く関連している。しかし、年齢構成は地域によって不均衡であるため、年齢が自殺のリスクの交絡要因となる。Rothman (2012) が検討しているように、期待死亡数の構成において交絡を統制するためには、年齢区分で層化することで交絡を統制し、その後地域 (層全体) での要約値を得るために標準化した期待死亡数を求めるという方法が考えられる。ここで、標準化は、加重平均

<sup>1</sup> 関西大学大学院心理学研究科

<sup>2</sup> SMR は、100 倍されている場合もあり、その場合は 100 が基準となる。

をとることにより区分別の率を単一の要約値へまとめる方法である。この加重平均に用いられる重みが、基準（全国）の自殺率であるため、基準の自殺率を対象地域の年齢分布に当てはめた場合の死亡数、つまり期待死亡数となる。この期待死亡数と実際に観測された死亡数との比をとることで **SMR** となる。

その他に、この指標を利用することの利点としては、自殺統計資料では特に入手困難ことが多い、地域ごとの年齢階級別死亡数を計算に必要としない点である。しかし、**SMR** は、人口が少ない地域において変動が大きい欠点がある。そのため、最も際立った特徴を示す地域の **SMR** は、単に信頼のないデータによるためであるかもしれない。有用な解釈を行うことは不可能となってしまう。推定に際しての標準誤差が結果として大きくなるということもできる。この欠点を補うために、たとえば、**Efron & Morris (1973)** が検討しているように、ベイズ推定の枠組みによる縮約推定量が用いられることがある。全体での枠組みを利用することで、個別の推定において全体の力による説得力の借用をする縮約推定量は、標本変動による標準誤差が大きい際に、より精度の高い推定値を得ることができる。このような縮約推定量は、**Stein (1956)**、**James & Stein (1961)** が正規分布において最尤推定量よりも優越することを示してから、様々な分野で応用がなされている（数理的な応用については久保川（2004）など）。例えば、経済学の分野での最適なポートフォリオの問題（**Jacquier & Polson, 2011**）、心理学の分野での個人の成長曲線の推定（**Singer & Willett, 2003**）、生物学でのコウチョウの寄生率の推定（**Link & Hahn, 1996**）などがある。特に、社会経済や疫学などの小地域推定において多く用いられている（**Maiti, 2005**）。

本稿の目的は、地域の自殺発生状況の実態を計量的に把握することに貢献することである。実態の把握を困難にしている要因の一つは、人口が少ない地域での自殺発生状況である。そこで、自殺発生状況に関する精度の高い指標を構成できるベイズ推定による方法を検討する。本稿では、ここで紹介してきた相対リスクに注目する。相対リスクは、自殺では、地域ごとの標準化死亡比として使用されている。その最尤推定量、経験ベイズ推定量、階層ベイズ推定量の3種類の推定量について検討することとする。したがって、**Clayton & Kaldor (1987)** などに従いながら、これらの推定量の構成と特長についてまとめ、相対リスクのベイズの推定方法の手順を紹介し、兵庫県における男性の自殺発生数を対象にして、相対リスクのベイズ推定量を求める際に前提となる仮定やその有効性を論じてみることにする。

## 2. 相対リスクの最尤推定量

本節では、相対リスクの推定量としての最尤推定量について検討する。まず、 $\theta_i$  を、地域  $i$  ( $i=1, \dots, m$ ) における、1年間の1人当たりの潜在的な自殺の相対リスクとする。準備として、ここでは、以下を想定しておくことにする。

- 1) 地域*i*は、それぞれ固有の自殺リスクを有している。
- 2) この自殺リスクに応じて、地域*i*は、それぞれ潜在的な自殺発生数を有している。
- 3) この自殺発生数と、全地域を総計した年齢別自殺率をもとに求められる地域*i*の自殺発生数の期待値の比を求めることができる。これを相対リスク $\theta_i$ とする<sup>3</sup>。
- 4) この $\theta_i$ は、地域*i*に固有の潜在的な相対リスクであり、直接観測することができない。そのため、実現数から推定値を計算する。
- 5) 推定量は、(1)式のようにして求められ、潜在的な自殺発生数にポワソン分布を仮定した場合<sup>4</sup>、この値は、 $\theta_i$ の最尤推定量となっている。つまり、

$$\hat{\theta}_i^{ML} = \frac{d_i}{e_i} \quad \dots\dots\dots (1)$$

である。

相対リスクの推定量の計算には、 $d_i$ は地域*i*における自殺発生数であり、 $e_i$ は

$$e_i = \sum_{h=1}^H n_{ih} P_h \quad \dots\dots\dots (2)$$

として求められる期待死亡数が選択されることが多い。ここで、 $n_{ih}$ は、地域*i*における年齢階級*h* ( $h=1, \dots, H$ )の人口、 $P_h$ は基準集団(全地域)における年齢階級*h*の自殺率である。

一般に、生起確率や割合などの比を推定する際には、比への要約によって分子と分母の独立した情報が失われ、また比の分布は扱いにくい問題点がある。そこで、(1)式から

$$d_i = e_i \times \theta_i \quad \dots\dots\dots (3)$$

としてモデル化を行う<sup>5</sup>。まず、地域*i*の自殺発生数 $d_i$ がポアソン分布

$d_i \overset{\text{ind}}{\sim} \text{Poisson}(e_i \theta_i)$ に従うと仮定する。この仮定の意味するところは、ポアソン分布

<sup>3</sup> 相対リスク $\theta_i$ の定義域は、 $0 \leq \theta_i \leq \infty$ である。

<sup>4</sup> 潜在的な自殺発生数にポワソン分布を仮定することは、自殺の発生という事象が、人口比で比較的小さいことを考えると、自然な仮定といえる。

<sup>5</sup> このように割算値を回避する方法は、一般化線形モデルの文脈ではオフセット (offset) 項と呼ばれる方法である (久保, 2012; Qian, 2010)。

の期待値と分散より自殺発生数の  $E[d_i] = e_i \theta_i$ ,  $\text{Var}[d_i] = e_i \theta_i$  であり, また, 潜在的な相対リスクは地域  $i$  で異なることを仮定している. この場合, 相対リスク  $\theta_i$  の確率密度関数は,

$$p(d_i) = \frac{(e_i \theta_i)^{d_i} \exp(-e_i \theta_i)}{d_i!}, \quad d_i = 0, 1, \dots, \theta_i > 0 \quad \dots\dots\dots (4)$$

である.

この  $\theta_i$  の最尤推定量を求める. この尤度関数は,

$$L(\theta_i) = \frac{(e_i \theta_i)^{d_i} \exp(-e_i \theta_i)}{d_i!} \quad \dots\dots\dots (5)$$

である<sup>6</sup>. したがって, 対数尤度関数は,

$$\ell(\theta_i) = d_i \log(e_i \theta_i) - e_i \theta_i - \log(d_i!) \quad \dots\dots\dots (6)$$

となる.  $\theta_i$  で偏微分して 0 とおくと,

$$\frac{\partial \ell(\theta_i)}{\partial \theta_i} = \frac{d_i e_i}{e_i \theta_i} - e_i = 0 \quad \dots\dots\dots (7)$$

となる. そのため,  $\theta_i$  の最尤推定量は,

$$\hat{\theta}_i^{ML} = \frac{d_i}{e_i} \quad \dots\dots\dots (8)$$

として求められ, (1)式が  $\theta_i$  の最尤推定量であることが示される<sup>7</sup>.

次に, この最尤推定値  $\hat{\theta}_i^{ML}$  の期待値と分散は,  $E[d_i] = e_i \theta_i$ ,  $\text{Var}[d_i] = e_i \theta_i$  を用いて,

$$E[\hat{\theta}_i^{ML}] = \frac{E[d_i]}{e_i} = \frac{e_i \theta_i}{e_i} = \theta_i \quad \dots\dots\dots (9)$$

<sup>6</sup> 尤度関数  $L(\theta_i)$  は, 代数的には確率密度関数  $p(d_i)$  と同じであるが, 表記が  $\theta_i$  を所与としたときの確率変数  $d_i$  を考えることから,  $d_i$  を所与としたときのパラメータ  $\theta_i$  と変更される.

<sup>7</sup> より厳密には, (11)式のように  $\ell(\theta_i)$  の 2 階導関数が負値となり, すべての局所最大値 (極値) の中で最大となる値が最尤推定量となる.

$$\text{Var}[\hat{\theta}_i^{ML}] = \frac{\text{Var}[d_i]}{e_i^2} = \frac{e_i \theta_i}{e_i^2} = \frac{\theta_i}{e_i} \quad \dots\dots\dots (10)$$

を得る。すなわち、 $\hat{\theta}_i^{ML}$  は  $\theta_i$  の不偏推定量 (unbiased estimator) である<sup>8</sup>。

次に、 $\text{Var}[\hat{\theta}_i^{ML}]$  がクラメール・ラオの下界 (Cramér-Rao's lower bound) に等しいかを検討してみることにする<sup>9</sup>。 $\ell(\theta_i)$  の 2 階導関数は、

$$\frac{\partial^2 \ell(\theta_i)}{\partial \theta_i^2} = -\frac{d_i}{\theta_i^2} \quad \dots\dots\dots (11)$$

となるから、フィッシャー情報量 (Fisher information) は、

$$I(\theta_i) = -E\left[\frac{\partial^2 \ell(\theta_i)}{\partial \theta_i^2}\right] = \frac{E[d_i]}{\theta_i^2} = \frac{e_i \theta_i}{\theta_i^2} = \frac{e_i}{\theta_i} \quad \dots\dots\dots (12)$$

となる。そのため、

$$I(\theta_i)^{-1} = \frac{\theta_i}{e_i} = \text{Var}[\hat{\theta}_i^{ML}] \quad \dots\dots\dots (13)$$

となり、 $\hat{\theta}_i^{ML}$  は  $\theta_i$  の有効推定量 (efficient estimator) である。したがって、 $\hat{\theta}_i^{ML}$  は、不偏性を有する推定量のクラスにおいて、一様に最小の分散を持つという意味で最適な推定量である。

### 3. 相対リスクのベイズ推定量

前節では、自殺発生数にポアソン分布を仮定した場合において、相対リスクの最尤推定量  $\hat{\theta}_i^{ML}$  は、有効性という推定量が持つべき望ましい性質を備えていることを示した。しかし、(10)式で示された通り、期待死亡数  $e_i$  が小さいほど、標準誤差は大きく

<sup>8</sup> 推定量  $\hat{\theta}_i$  が、 $E(\hat{\theta}_i) = \theta_i$  という性質をもつ場合に不偏推定量と呼ばれる。これは、繰り返し標本を抽出してそのたびに推定値を得ると、その期待値が  $\theta_i$  になることを示す。

<sup>9</sup> 不偏推定量において、それ以上分散を小さくできないという分散の限界を与えるのが、クラメール・ラオの下界であり、この下界を達する不偏推定量を有効推定量または一様最小分散不偏推定量 (uniformly minimum variance unbiased estimator) という。フィッシャー情報量  $I(\theta_i)$  が正であり、微分と積分の交換が保障されるなどの正則条件のもとで、 $\text{Var}(\hat{\theta}_i) \geq I(\theta_i)^{-1}$  が成立する (竹村, 1991)。そのため、この下界はフィッシャー情報量の逆数として与えられる。

なる性質をもつ。言い換えれば、最尤推定量  $\hat{\theta}_i^{ML}$  は、人口が少ない小地域において実際の死亡数  $d_i$  の微小な変動に対しても比較的大きな揺らぎを生じる。そこで、以下では、この問題を回避するために、 $\theta_i$  に事前分布を仮定するベイズ推定の枠組みで縮約推定量 (shrinkage estimator) を構成する。なお、Clayton & Kaldor (1987), Datta, Ghosh, & Waller (2000), Rao (2003), 丹後・横山・高橋 (2007) などを参考にしながら、相対リスクの基本となる Poisson-Gamma モデルを紹介してみることにする。

Poisson-Gamma モデルは、まず前と同様に、地域  $i$  の自殺発生数  $d_i$  はポアソン分布  $d_i \stackrel{\text{iid}}{\sim} \text{Poisson}(e_i \theta_i)$  に従うと仮定する。ただし、 $\theta_i$  は確率変数として扱い事前分布を想定するため、相対リスク  $\theta_i$  は以下の条件付密度関数となる。

$$p(d_i | \theta_i) = \frac{(e_i \theta_i)^{d_i} \exp(-e_i \theta_i)}{d_i!}, \quad d_i = 0, 1, \dots, \theta_i > 0 \quad \dots\dots\dots (14)$$

また、自殺発生数  $d_1, \dots, d_m$  は相対リスク  $\theta_1, \dots, \theta_m$  が与えられたもとで互いに独立であり、空間的に近接する地域間でも自殺発生数の間には相関がないと仮定する。つまり、

$$p(d_1, \dots, d_m | \theta_1, \dots, \theta_m) = \prod_{i=1}^m p(d_i | \theta_i) \quad \dots\dots\dots (15)$$

である。

次に、潜在的な相対リスク  $\theta_i$  については、ガンマ分布を想定し、 $\theta_i$  は独立同一分布<sup>10</sup> するものとする。つまり、 $\theta_i \stackrel{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta)$  である。これは、相対リスクは地域ごとで異なりうるが、地域全体  $\{\theta_1, \dots, \theta_m\}$  としては平均  $\alpha/\beta$ 、分散  $\alpha/\beta^2$  であることを仮定している<sup>11</sup>。ポアソン分布のパラメータ  $\theta_i$  の事前分布としてガンマ分布を仮定する

<sup>10</sup> 確率変数  $X_1, \dots, X_n$  が、互いに独立かつ同じ分布にしたがっている時、 $X_1, \dots, X_n$  は独立同一分布 (Independently and Identically Distributed) であるという。

<sup>11</sup>  $\text{Gamma}(\alpha, \beta)$  の  $\alpha$  は形状母数 (shape)、 $\beta$  は逆尺度母数 (rate) とよばれる未知のパラメータ

のは、ガンマ分布が、ポワソン分布の自然共役事前分布となっており、そのため事後分布もガンマ分布し、計算の利便性があること、また非負の値をとって単峰の分布への近似がよいことが理由となっている<sup>12</sup>。確率密度関数は、

$$g(\theta_i; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\beta\theta_i), \quad \theta_i > 0, \alpha > 0, \beta > 0 \quad \dots\dots\dots (16)$$

となる。

ところで、ベイズの定理より、事後分布は、

$$\pi(\theta_1, \dots, \theta_m | d_1, \dots, d_m) = \frac{\pi(d_1, \dots, d_m | \theta_1, \dots, \theta_m) \pi(\theta_1, \dots, \theta_m)}{\int_{\Theta} \pi(d_1, \dots, d_m | \theta_1, \dots, \theta_m) \pi(\theta_1, \dots, \theta_m) d\theta} \quad \dots\dots\dots (17)$$

となる。ただし、 $d\theta = d\theta_1 \cdots d\theta_m$ とする。ここで、(15)式の自殺発生数の条件付独立と相対リスクの独立より、

$$\pi(\theta_1, \dots, \theta_m | d_1, \dots, d_m) = \frac{\prod_{i=1}^m p(d_i | \theta_i) \prod_{i=1}^m g(\theta_i; \alpha, \beta)}{\int_{\Theta} \left\{ \prod_{i=1}^m p(d_i | \theta_i) \prod_{i=1}^m g(\theta_i; \alpha, \beta) \right\} d\theta} \quad \dots\dots\dots (18)$$

$$= \prod_{i=1}^m \left\{ \frac{p(d_i | \theta_i) g(\theta_i; \alpha, \beta)}{\int_{\Theta_i} p(d_i | \theta_i) g(\theta_i; \alpha, \beta) d\theta_i} \right\} \quad \dots\dots\dots (19)$$

となる。つまり、事後分布は、

$$\pi(\theta_1, \dots, \theta_m | d_1, \dots, d_m) = \prod_{i=1}^m \pi(\theta_i | d_i; \alpha, \beta) \quad \dots\dots\dots (20)$$

という関係にあり、事後分布は地域*i*ごとで考えればよいことがわかる。尤度  $p(d_i | \theta_i)$

と事前分布  $g(\theta_i; \alpha, \beta)$  から事後分布は、

$$\pi(\theta_i | d_i; \alpha, \beta) = \frac{p(d_i | \theta_i) g(\theta_i; \alpha, \beta)}{\int_{\Theta_i} p(d_i | \theta_i) g(\theta_i; \alpha, \beta) d\theta_i} \quad \dots\dots\dots (21)$$

であり、平均は  $\alpha / \beta$ 、分散は  $\alpha / \beta^2$  である。

<sup>12</sup> ベイズ統計学では、ある尤度関数に対して事前分布と事後分布が同じ分布族に属する場合、共役事前分布 (conjugate prior) と呼ばれる。

$$= \frac{(\beta + e_i)^{\alpha + d_i}}{\Gamma(\alpha + d_i)} \theta_i^{\alpha + d_i - 1} \exp\{-(\beta + e_i)\theta_i\} \cdots \cdots \quad (22)$$

となる<sup>13</sup>。すなわち、事後分布は $\theta_i | d_i \stackrel{\text{ind}}{\sim} \text{Gamma}(\alpha + d_i, \beta + e_i)$ である。潜在的な相対リスク $\theta_i$ のベイズ推定量は、事後分布の期待値として、

$$\hat{\theta}_i^B = E[\theta_i | d_i] = \frac{\alpha + d_i}{\beta + e_i} \cdots \cdots \quad (23)$$

として得られる<sup>14</sup>。

#### 4. 縮約推定量としてのベイズ推定量

さて、ベイズ推定量はどのような特徴を持つのであろうか。詳細は Efron & Morris (1973), Morris (1983) や久保川 (2004) を参照してもらうことにして、ここでは概要だけを述べる。一般に、望ましい推定量の基準として、

$$R(\theta, \hat{\theta}) = E\left\{\left(\hat{\theta}_i - \theta_i\right)^2\right\} \cdots \cdots \quad (24)$$

がある。これは、平均二乗誤差 (mean square error) とよばれる。この基準として選択されている推定量が、先に示した最尤推定量ということになる。同じく、この基準により、事後分布から点推定量を決めた推定量は、(23)式のベイズ推定量でもある。

ここで、平均二乗誤差を基準とするならば、個別に $\theta_i$ を推定する問題を考えたものが、先に示したように、最尤推定量よりも一様に平均二乗誤差が小さい推定量は存在しないことになる。しかし、地域 $i$ が複数である時に、 $\{\theta_1, \dots, \theta_m\}$ を同時に推定することを考え、平均二乗誤差の和

$$R(\theta, \hat{\theta}) = \sum_{i=1}^m \left[ E\left\{\left(\hat{\theta}_i - \theta_i\right)^2\right\} \right] \cdots \cdots \quad (25)$$

を基準として考えると、このとき、ベイズ推定量が最尤推定量を改良することがわかっている<sup>15</sup>。つまり、ベイズ推定量は全体としての平均二乗誤差が最尤推定量よりも小さい推定量である。ただし、個々の推定値においては、必ずしも誤差が小さくなっ

<sup>13</sup> (21)式から(22)式への変換は、付録2参照。

<sup>14</sup> ここで、事後分布の期待値を点推定値とする決定は、統計的決定理論あるいは意思決定の問題に関わる (Lancaster, 2004; 竹村, 1991)。事後分布から、様々な点推定値を選択できることがベイズ統計の特徴の一つであるが、一般的に用いられる平均二乗誤差を基準による点推定値は、事後分布の期待値である (松原, 2010)。

<sup>15</sup> Stein (1956), James & Stein (1961) により示され、個々に推定する問題で最尤推定量が許容的であるにもかかわらず、3以上では非許容的となる (久保川, 2004)。



ているとは限らないことに注意が必要である。

次に、このベイズ推定量は、実際の死亡数  $d_i$  の微小な変動に対しても安定的になる。まず、(23)式より、

$$\hat{\theta}_i^B = \left(1 - \frac{\beta}{\beta + e_i}\right) \frac{d_i}{e_i} + \left(\frac{\beta}{\beta + e_i}\right) \frac{\alpha}{\beta} \quad \dots\dots\dots (26)$$

と変形できる。ここで、 $d_i/e_i = \hat{\theta}_i^{ML}$  であり、また  $\alpha/\beta$  は事前分布  $\theta_i$  の平均となっており。すなわち、相対リスクのベイズ推定量  $\hat{\theta}_i^B$  は、地域  $i$  における相対リスクの最尤推定量と、事前分布の平均との加重平均である。先に説明したように、事前分布の平均は相対リスクの地域全体での平均を表す。このため、ベイズ推定量は、標本での推定量を全体の推定量に縮約しており、個別の推定において全体の力による説得力の借用をしている (borrowing strength from the ensemble)。このような推定量は、縮約推定量と呼ばれる。

この縮約推定量の特長は、加重部分に注目すると、ベイズ推定値  $\hat{\theta}_i^B$  は、期待度数  $e_i$  が多く標準誤差が小さい地域では、その地域  $i$  での相対リスクの最尤推定値に近づき、逆に期待度数  $e_i$  が少なく標準誤差が大きい地域では、相対リスクの地域全体の平均に近づくことがわかる。期待度数  $e_i$  の大小によって地域  $i$  の相対リスクを修正していることを示す。

## 5. 経験ベイズ法による相対リスクの推定

前節までは、ベイズ推定量  $\hat{\theta}_i^B$  が縮約推定量であることを示し、相対リスクのベイズ推定量は、ベイズの定理に基づく事後分布の期待値として求められることを説明した。ベイズ推定では、いったん事前分布を与えられれば、事後分布はベイズの定理によりいわば機械的に計算できる。問題は、事前分布の未知であるパラメータ  $\alpha$ 、 $\beta$  である。本来、ベイズ推定では、事前分布は、観測値を得る以前に利用可能な情報を表すもので、事前分布とそのパラメータはデータが得られる前に与えられる。これに対して、パラメータ  $\alpha$ 、 $\beta$  を確率変数ではない未知定数として扱い、経験的に与えるのが経験ベイズ (Empirical Bayes ; EB) 法である。経験的に与えるとは、データを利用して事前分布のパラメータを推定することを意味する。このような理由により、経験ベ

ズ法は、次節で示す完全なベイズ推定である階層ベイズ法の近似としてみなされる。本節では、経験ベイズ法について取り上げる。

まず、地域  $i$  の事後分布の密度関数は、(21)式より、

$$\pi(\theta_i | d_i; \alpha, \beta) = \frac{p(d_i | \theta_i) g(\theta_i; \alpha, \beta)}{\int_{\Theta_i} p(d_i | \theta_i) g(\theta_i; \alpha, \beta) d\theta_i} \quad \dots\dots\dots (27)$$

であるが、密度関数の核 (kernel) は、分母の規格化定数を省略した

$$\pi(\theta_i | d_i; \alpha, \beta) \propto p(d_i | \theta_i) g(\theta_i; \alpha, \beta) = p(d_i, \theta_i; \alpha, \beta) \quad \dots\dots\dots (28)$$

となり、事後分布は同時分布に比例する。  $\propto$  は比例することを表す。ここで、右辺において、  $\theta_i$  はここでは推論の対象とならない局外パラメータ (nuisance parameter) として積分消去することで、

$$p(d_i; \alpha, \beta) = \int_{\Theta_i} p(d_i | \theta_i) g(\theta_i; \alpha, \beta) d\theta_i \quad \dots\dots\dots (29)$$

を得る。これは、パラメータ  $\alpha$  ,  $\beta$  のときの標本  $d_i$  の確率密度関数であるので、すべ

ての標本  $\{d_1, \dots, d_m\}$  を所与としたときのパラメータの尤度関数は、

$$L(\alpha, \beta) = \prod_{i=1}^m p(d_i; \alpha, \beta) \quad \dots\dots\dots (30)$$

$$= \prod_{i=1}^m \left\{ \int_{\Theta_i} p(d_i | \theta_i) g(\theta_i; \alpha, \beta) d\theta_i \right\} \quad \dots\dots\dots (31)$$

と考えられる。この尤度関数は、事後分布に比例する尤度関数を周辺化した周辺尤度 (marginal likelihood) と呼ばれる。そして、経験ベイズ法では、この周辺尤度  $L(\alpha, \beta)$  を最大にするようにパラメータを推定する。つまり、

$$\hat{\alpha}^{EB} = \arg \max_{\alpha} L(\alpha, \beta), \quad \hat{\beta}^{EB} = \arg \max_{\beta} L(\alpha, \beta) \quad \dots\dots\dots (32)$$

である<sup>16</sup>。このように、尤度関数を最大にする未知定数のパラメータを推定するので、推定の形式としては最尤推定法が用いられる。

ところで、(29)式は、

---

<sup>16</sup> 記号「 $\arg \max$ 」は、argument of the maximum のことであり、最大値をとる引数の値を返すことを示す。

$$p(d_i; \alpha, \beta) = \binom{d_i + \alpha - 1}{d_i} \left( \frac{\beta}{e_i + \beta} \right)^\alpha \left( \frac{e_i}{e_i + \beta} \right)^{d_i} \quad \dots\dots\dots (33)$$

となる<sup>17</sup>. これは  $d_i \stackrel{\text{iid}}{\sim} \text{Negative Binomial}(\alpha, \beta)$  である. これにより, 周辺尤度の尤度関数は,

$$L(\alpha, \beta) = \prod_{i=1}^m \left\{ \binom{d_i + \alpha - 1}{d_i} \left( \frac{\beta}{e_i + \beta} \right)^\alpha \left( \frac{e_i}{e_i + \beta} \right)^{d_i} \right\} \quad \dots\dots\dots (34)$$

となり, したがって, 対数尤度関数は,

$$\ell(\alpha, \beta) = \sum_{i=1}^m \left\{ \sum_{h=0}^{d_i-1} \log(\alpha + h) + \alpha \log \beta - (d_i + \alpha) \log(e_i + \beta) + d_i \log(e_i) \right\} \dots (35)$$

となる. 最尤推定量を求めるため,  $\alpha$  と  $\beta$  でそれぞれ偏微分して,

$$\frac{\partial \ell(\alpha, \beta)}{\partial \alpha} = \sum_{i=1}^m \sum_{h=0}^{d_i-1} \frac{1}{\alpha + h} + m \log \beta - \sum_{i=1}^m \log(e_i + \beta) \quad \dots\dots\dots (36)$$

$$\frac{\partial \ell(\alpha, \beta)}{\partial \beta} = \frac{m\alpha}{\beta} - \sum_{i=1}^m \frac{d_i + \alpha}{e_i + \beta} \quad \dots\dots\dots (37)$$

より, 上式の右边=0 と置いて,

$$\sum_{i=1}^m \sum_{h=0}^{d_i-1} \frac{1}{\alpha + h} + m \log \beta - \sum_{i=1}^m \log(e_i + \beta) = 0 \quad \dots\dots\dots (38)$$

$$\frac{\alpha}{\beta} = \frac{1}{m} \sum_{i=1}^m \frac{d_i + \alpha}{e_i + \beta} \quad \dots\dots\dots (39)$$

となるが, (38)式と(39)式から  $\hat{\alpha}^{EB}$  と  $\hat{\beta}^{EB}$  の陽表解は一般に得られない (Clayton & Kaldor, 1987). そのため, 経験ベイズ法の最尤推定値は, 繰り返しの数値計算によって求める. 一方で, 簡便には, モーメント推定を適用し,  $\hat{\alpha}$  と  $\hat{\beta}$  を求めることもできる (Marshall, 1991; Rao, 2003). まず,

$$e. = \frac{1}{m} \sum_{i=1}^m e_i \quad \dots\dots\dots (40)$$

---

<sup>17</sup> (29)式から(33)式への変換は, 付録 1 を参照.

とする．加重の標本平均

$$\hat{\theta}_{e\cdot} = \frac{1}{m} \sum_{i=1}^m \left( \frac{e_i}{e} \right) \hat{\theta}_i \quad \dots\dots\dots (41)$$

と加重の標本分散

$$s_e^2 = \frac{1}{m} \sum_{i=1}^m \left( \frac{e_i}{e} \right) (\hat{\theta}_i - \hat{\theta}_{e\cdot})^2 \quad \dots\dots\dots (42)$$

から，これらの期待値の方程式を作り， $\hat{\alpha}$  と  $\hat{\beta}$  についてのモーメント方程式を解くこと

とで，モーメント推定量は，

$$\frac{\hat{\alpha}}{\hat{\beta}} = \hat{\theta}_{e\cdot} \quad \dots\dots\dots (43)$$

$$\frac{\hat{\alpha}}{\hat{\beta}^2} = s_e^2 - \frac{\hat{\theta}_{e\cdot}}{e} \quad \dots\dots\dots (44)$$

と求められる．このモーメント推定値は，最尤推定値における数値計算の初期値としても使用できる．

## 6. 階層ベイズ法による相対リスクの推定

経験ベイズ法では，事前分布のパラメータ  $\alpha$ ， $\beta$  を未知定数として，データから推定することを行った．ベイズ統計と頻度論をいわば折衷的に使用していたが，本節では，完全ベイズ法（Full Bayes approach）ともよばれる階層ベイズ法（Hierarchical Bayes approach ; HB）による相対リスクの推定について考える．しかし，階層ベイズ法では，一部のパラメータに関する周辺事後分布や期待値を求める際には高次元の積分が必要となり，事後分布は多次元になることが多く解析的な方法や古典的なモンテカルロ法では一般に計算が困難になる（たとえば，伊庭（2005）など）．そこで，階層ベイズ法では，解析的に求まらない事後分布からのサンプリングを可能にするマルコフ連鎖モンテカルロ法によるサンプリングを行う．6-1 節では，本稿で使用するマルコフ連鎖モンテカルロ法の概要について説明する．6-2 節で，Poisson-Gamma モデルの階層ベイズ法による相対リスクの推定について説明する．

### 6-1 マルコフ連鎖モンテカルロ法の概要

マルコフ連鎖モンテカルロ（Markov chain Monte Carlo ; MCMC）法は，マルコフ連鎖の定常分布から目標分布（ここでは事後分布）と一致するサンプルを発生させる方法であり，通常のランダムサンプリング法とは異なり，1 回前にサンプリングさ

れた値に依存させて次の値をサンプリングする方法の総称である。MCMC法について、大森 (2001), 伊庭 (2005), 和合 (2005), Gelman, Carlin, Stern, Dunson, Vehtari, & Rubin (2013), Geweke, Koop, & van Dijk (2011), Robert & Casella (2010) を参考として、本節では、MCMC法の最も一般的に使用されるギブス・サンプラーとメトロポリス - ヘイスティングス・アルゴリズムを概説する<sup>18</sup>。

代表的なMCMC法に、ギブス・サンプラー (Gibbs sampler) がある。ギブス・サンプラーの考え方は、高次元の問題を次元の小さな複数の問題に分解して効率的に解くことである。高次元の事後分布  $\pi(\boldsymbol{\theta} | \mathbf{y})$  を解析的に求められない場合であっても、未知の多次元パラメータ  $\boldsymbol{\theta}$  をいくつかのブロック  $\{\theta_1, \dots, \theta_k\}$  に分けた完全条件付分布

$\pi(\theta_i | \{\theta_j\}_{j \neq i}, \mathbf{y})$  ( $i = 1, \dots, k$ ) は解析的に求められる場合がある。ここで、 $\mathbf{y}$  はデータを

表し、完全条件付分布とは、データを含むモデルの他の全ての確率変数が与えられた時のその量の分布である。こうした場合に用いられるのがギブス・サンプラーである。

適当な初期値  $\{\theta_1^{(0)}, \dots, \theta_k^{(0)}\}$  からスタートする (右上の添え字はループ数を表す)。ま

ず、完全条件付分布  $\pi(\theta_1 | \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, \mathbf{y})$  から  $\theta_1^{(1)}$  をサンプリングして、次に、

$\pi(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)}, \mathbf{y})$  から  $\theta_2^{(1)}$  をサンプリングする。これを繰り返し、最後に

$\pi(\theta_k | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{k-1}^{(1)}, \mathbf{y})$  から  $\theta_k^{(1)}$  をサンプリングする。以上を第 1 ループとする。次

に、この第 1 ループでサンプリングされた  $\{\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_k^{(1)}\}$  からスタートして、同様

に第 2 ループを行い、 $\{\theta_1^{(2)}, \theta_2^{(2)}, \dots, \theta_k^{(2)}\}$  をサンプリングする。これを繰り返すと、第

$\ell$  ループでは  $\{\theta_1^{(\ell)}, \theta_2^{(\ell)}, \dots, \theta_k^{(\ell)}\}$  がサンプリングされることになる。このように、高次

元の問題であっても、シミュレーションはすべて単変量になる。適切な条件のもとで、

$\ell \rightarrow \infty$  とすると、事後分布  $\pi(\theta_1, \theta_2, \dots, \theta_k | \mathbf{y})$  に収束することが知られている<sup>19</sup>。そこ

で、初期値に依存する  $M$  ループまでにサンプリングされた値を捨て<sup>20</sup>、その後の  $N$  ル

ープを行ってサンプリングされた  $\{\theta_1^{(\ell)}, \theta_2^{(\ell)}, \dots, \theta_k^{(\ell)}\}$  ( $\ell = M + 1, M + 2, \dots, M + N$ )

<sup>18</sup> ギブス・サンプラーは、メトロポリス - ヘイスティングス・アルゴリズムの特別な場合である。

<sup>19</sup> マルコフ連鎖に必要な条件は、万遍なくサンプリングするための条件であり、既約性、正再帰性、非周期性と呼ばれる (大森, 2001)。

<sup>20</sup> 初期値に影響を受けて捨てられる期間をバーンイン (burn-in) 期間と呼ばれる。

は、事後分布  $\pi(\theta_1, \theta_2, \dots, \theta_k | \mathbf{y})$  からサンプリングされた値とみなすことができる。

しかし、実際にはギブス・サンプラーでの完全条件付分布が解析的に得られない場合がある。その場合に使用されるのが、メトロポリス - ヘイスティングス (Metropolis-Hastings ; M-H) アルゴリズムである。これは、確率分布  $\pi(x)$  からサンプリングしたいが、特殊な分布であるため直接はサンプリングできない場合に用いられる。ここで、 $\pi(x)$  は目標分布と呼ばれ、規格化定数以外の部分は解析的に求められているとする。M-H アルゴリズムでは、 $\pi(x)$  を近似した容易にサンプリングができる提案分布  $h(x)$  と初期値を決め、サンプリングしたい目標分布と提案分布の違いを、定常分布として推移する条件を満たすようにする操作を含めることで、目標分布からのサンプリングを可能とするアルゴリズムである。この修正する操作は、候補の値をマルコフ連鎖の次の値として受け入れるかを判断する受理確率の計算で構成される。

提案分布  $h(x)$  を選択する一つの自然な方法は、先にサンプリングした値を検討し、続く値を生成することで、マルコフ連鎖の現在値の周辺を探索することである。最も簡単な方法としては、原点を中心にして対称な密度である平均  $0$  で分散  $\sigma^2$  の正規分布に従う  $z$  を用いて、 $\theta^* = \theta^{(j-1)} + z$  を候補とする。そのため、候補の値は、正規分布を提案分布とする  $\theta^* | \theta^{(j-1)} \sim Normal(\theta^{(j-1)}, \sigma^2)$  となる<sup>21</sup>。これは、ランダムウォーク連鎖によるM-Hアルゴリズムと呼ばれ、次の手順でサンプリングする。まず、初期値  $\theta^{(0)}$  と尺度パラメータ  $\sigma^2$  を決め、①提案分布  $Normal(\theta^{(j-1)}, \sigma^2)$  から候補の値  $\theta^*$  をサンプリングする。②受理確率  $P(\theta^{(j-1)}, \theta^*) = \min[R, 1]$  を計算する。ここで、 $R = \pi(\theta^*) / \pi(\theta^{(j-1)})$  である。③受理確率  $P(\theta^{(j-1)}, \theta^*)$  で  $\theta^{(j)} = \theta^*$  とする。受理されない場合は  $\theta^{(j)} = \theta^{(j-1)}$  とする<sup>22</sup>。以上を繰り返すと、 $\theta^{(j)}$  の分布は適切な条件の下で  $j \rightarrow \infty$  のときに目標分布に収束することが知られている。本稿では、ギブス内メトロポリス (Metropolis within Gibbs) として、ギブス・サンプラーで事後分布を単変量の完全条件付分布に分解してサンプリングする。その際、完全条件付分布からの直接サンプリングが困難な場合には、ランダムウォーク連鎖によるM-Hアルゴリズムを使用する。

MCMC 法は、理論的にはマルコフ連鎖が定常分布に収束することが示されるが、実践面ではMCMC アルゴリズムをいつ打ち切るかに明確な基準はなく、MCMC サンプルの出力から収束を事後的にチェックして試行錯誤することとなる。収束の診断は、最も基本的にはマルコフ連鎖のトレースプロット (標本経路図) を描き、サンプルが初期値への依存、逸脱や非定常的な挙動を示すかを確認することである。また、標本

<sup>21</sup> 正規分布の分散  $\sigma^2$  は、尺度パラメータと呼ばれ、受理率により調整することとなる。

<sup>22</sup> 実装においては、一様分布を用いて採用の判定を行う。

自己相関関数のプロットから、自己相関が低くパラメータ空間全体を効率的に探索していることを確認する。収束に関する統計量としては、連鎖の分布がステップの推移で変化しないことを確認する指標である Geweke (1992) の統計量が用いられる。加えて、実験的に初期値を変えた複数のマルコフ連鎖を生成し、複数の連鎖が同一の領域をカバーするかを判断する Gelman & Rubin (1992) の  $\hat{R}$  指数がしばしば使用される。収束診断には、様々な方法があるが絶対的なものはない。この点に関しては、Robert & Casella (2010)、久保 (2012) が詳しい。

実際の MCMC 法の実行においては、完全条件付分布などの MCMC 法における詳細な設定を考えなくても、モデルを指定すれば、MCMC 法の汎用計算プログラムである Stan (Stan Development Team, 2014)、JAGS (Plummer, 2013)、BUGS などで計算することが可能である。しかし、その場合であっても、MCMC サンプルの収束診断を行う必要がある。そして、初期値が影響している一定のバーンイン期間を除き、抽出した MCMC サンプルを用いて、それぞれの地域における事後分布の期待値  $\hat{\theta}_i^\beta = E[\theta_i | d_i]$  を得る。

## 6-2 Poisson-Gamma モデルの階層ベイズ推定

さて、階層ベイズ法では、経験ベイズとは異なり、事前分布のパラメータ  $\alpha$ 、 $\beta$  についても定数ではなく確率変数として、さらなる事前分布である  $\pi(\alpha, \beta)$  を想定する<sup>23</sup>。したがって、階層ベイズ推定では、相対リスク  $\theta_1, \dots, \theta_m$  は、単に独立ではなくパラメータ  $\alpha$ 、 $\beta$  が与えられたもとで互いに独立である。つまり、

$$\pi(\theta_1, \dots, \theta_m | \alpha, \beta) = \prod_{i=1}^m \pi(\theta_i | \alpha, \beta) \quad \dots\dots\dots (45)$$

である。このように、地域  $i$  の事前分布は、パラメータ  $\alpha$ 、 $\beta$  の条件付分布として  $\pi(\theta_i | \alpha, \beta)$  と表現される。

パラメータ  $\alpha$ 、 $\beta$  の事前分布  $\pi(\alpha, \beta)$  について、事前に何らかの情報があれば、その情報を反映した事前分布を設定することができる。しかし、事前の情報がない場合、またできるだけ主観を排除するために、無情報事前分布 (noninformative prior distribution) がしばしば用いられる。Poisson-Gamma モデルの無情報事前分布としては、次のような提案がなされてきている。George, Makov, & Smith (1993) は、 $\alpha \sim \text{Exp}(b=1)$ 、 $\beta \sim \text{Gamma}(a=0.1, b=1)$  と事前分布を設定して、事後分布が正則となる方法を提案している。Natarajan, Ghosh, & Maiti (1998) は、事前分布を選択す

<sup>23</sup> これは、事前分布と区別して、超事前分布 (hyper prior) とも呼ばれる。また、1 段目の  $\pi(\theta | \alpha, \beta)$  だけでなく、2 段目の  $\pi(\alpha, \beta)$  も含めた  $\pi(\theta | \alpha, \beta) \pi(\alpha, \beta)$  を階層事前分布と呼ぶ場合もある。

る方法として広く使用されているジェフリーズ(Jeffreys)のルールに従って、 $\alpha \propto 1/\alpha$ ,  $\beta \sim \text{Gamma}(a = 0.5, b > 0)$ として<sup>24</sup>, 少なくとも1つは $d_i$ が0でない時に事後分布が正則となる事前分布を提案している. また, Lawson (2013) は, 平均10で分散100となるように,  $\alpha \sim \text{Exp}(b = 0.1)$ ,  $\beta \sim \text{Exp}(b = 0.1)$ としている. このように, 事前分布の選択には, 絶対的なものではなく, 事前分布の選択を変えて相対リスクのベイズ推定への影響を検討する感度分析が必要となる (Lancaster, 2004; Albert, 2009). ここでは, 事前に情報がないとして, George, Makov, & Smith (1993) の事前分布を採用することとして展開する. したがって, 階層ベイズ法でのPoisson-Gammaモデルは,

$$\text{I) } d_i \overset{\text{ind}}{\sim} \text{Poisson}(e_i \theta_i) \quad \dots\dots\dots (46)$$

$$\text{II) } \theta_i \sim \text{Gamma}(\alpha, \beta) \quad \dots\dots\dots (47)$$

$$\text{III) } \alpha \sim \text{Exp}(b_\alpha = 1), \quad \beta \sim \text{Gamma}(a = 0.1, b_\beta = 1) \quad \dots\dots\dots (48)$$

とする.

Poisson-Gamma モデルの事後分布を求めるため, MCMC 法のギブス・サンプラーによるサンプルを抽出するための完全条件付分布を決定する. 完全条件付分布を決定するための標準的なアプローチは, 完全条件付分布が考慮中のデータとすべての確率変数の同時分布に比例することにもとづく. ここで,  $\pi$  は確率密度関数,

$d_{1:m} = \{d_1, \dots, d_m\}$ ,  $\theta_{1:m} = \{\theta_1, \dots, \theta_m\}$ , また  $\theta_{-i}$  は  $\theta_{1:m}$  から  $\theta_i$  を除いた  $\{\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_m\}$  を示すこととして, 本モデルでのデータとすべての確率変数の同時分布は,

$$\pi(d_{1:m}, \theta_{1:m}, \alpha, \beta) = \pi(d_{1:m} | \theta_{1:m}, \alpha, \beta) \pi(\theta_{1:m} | \alpha, \beta) \pi(\alpha, \beta) \quad \dots\dots\dots (49)$$

$$= \pi(d_{1:m} | \theta_{1:m}) \pi(\theta_{1:m} | \alpha, \beta) \pi(\alpha, \beta) \quad \dots\dots\dots (50)$$

$$= \prod_{i=1}^m \pi(d_i | \theta_i) \prod_{i=1}^m \pi(\theta_i | \alpha, \beta) \pi(\alpha) \pi(\beta) \quad \dots\dots\dots (51)$$

となる. 展開の1行目は同時分布を条件付分布に分解しており, 2行目は $d_{1:m}$ は $\alpha$ と $\beta$ に依存しないためである. また, 3行目は, モデルの設定の(15)式と(45)式から $d_i$ と $\theta_i$ はそれぞれパラメータを与えられたもとで条件付独立であるためである.

ここで, i)  $\theta_i$  の完全条件付分布は,

---

<sup>24</sup> 記号  $\propto$  は, 比例関係をあらわす.



$$\pi(\theta_i | d_{1:m}, \theta_{-i}, \alpha, \beta) = \frac{\pi(d_{1:m}, \theta_{1:m}, \alpha, \beta)}{\pi(d_{1:m}, \theta_{-i}, \alpha, \beta)} \propto \pi(d_{1:m}, \theta_{1:m}, \alpha, \beta) \quad \dots\dots\dots (52)$$

$$\propto \prod_{i=1}^m \pi(d_i | \theta_i) \prod_{i=1}^m \pi(\theta_i | \alpha, \beta) \pi(\alpha) \pi(\beta) \quad \dots\dots\dots (53)$$

$$\propto \pi(d_i | \theta_i) \pi(\theta_i | \alpha, \beta) \quad \dots\dots\dots (54)$$

となる．展開の 1 行目は，分母の  $\pi(d_{1:m}, \theta_{-i}, \alpha, \beta)$  は  $\theta_i$  に関係せず定数とみなせるため， $\theta_i$  の完全条件付分布は同時分布に比例することとなる．これ以降も  $\theta_i$  に関する部分のみを残していく．モデルの設定より， $\pi(d_i | \theta_i)$  は Poisson 分布， $\pi(\theta_i | \alpha, \beta)$  は Gamma 分布であるので，上式は，

$$\propto \frac{(e_i \theta_i)^{d_i} \exp(-e_i \theta_i)}{d_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\beta \theta_i) \quad \dots\dots\dots (55)$$

$$\propto \theta_i^{d_i} \exp(-e_i \theta_i) \theta_i^{\alpha-1} \exp(-\beta \theta_i) \quad \dots\dots\dots (56)$$

$$\propto \theta_i^{d_i + \alpha - 1} \exp\{- (e_i + \beta) \theta_i\} \quad \dots\dots\dots (57)$$

となる．したがって， $\theta_i$  の完全条件付分布は，

$$\theta_i | d_{1:m}, \theta_{-i}, \alpha, \beta \overset{\text{ind}}{\sim} \text{Gamma}(d_i + \alpha, e_i + \beta) \quad \dots\dots\dots (58)$$

である． $m$  個の  $\theta_1, \dots, \theta_m$  は，それぞれ対応する死亡発生数  $d_i$  と期待死亡数  $e_i$  から計算されるこの完全条件付分布から抽出できる．

次に，ii)  $\alpha$  の完全条件付分布は，

$$\pi(\alpha | d_{1:m}, \theta_{1:m}, \beta) = \frac{\pi(d_{1:m}, \theta_{1:m}, \alpha, \beta)}{\pi(d_{1:m}, \theta_{1:m}, \beta)} \propto \pi(d_{1:m}, \theta_{1:m}, \alpha, \beta) \quad \dots\dots\dots (59)$$

$$\propto \prod_{i=1}^m \pi(d_i | \theta_i) \prod_{i=1}^m \pi(\theta_i | \alpha, \beta) \pi(\alpha) \pi(\beta) \quad \dots\dots\dots (60)$$

$$\propto \prod_{i=1}^m \pi(\theta_i | \alpha, \beta) \pi(\alpha) \quad \dots\dots\dots (61)$$

となり，モデルの設定から， $\pi(\theta_i | \alpha, \beta)$  は Gamma 分布， $\pi(\alpha)$  は Exp 分布であるので，上式は，

$$\propto \prod_{i=1}^m \left\{ \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\beta\theta_i) \right\} b_\alpha \exp(-b_\alpha \alpha) \quad \dots\dots\dots (62)$$

$$\propto \frac{\beta^{m\alpha}}{\Gamma^m(\alpha)} \prod_{i=1}^m (\theta_i^{\alpha-1}) \exp(-b_\alpha \alpha) \quad \dots\dots\dots (63)$$

となる。これは、標準的な分布の形ではない。

最後に、iii)  $\beta$  の完全条件付分布は、

$$\pi(\beta | d_{1:m}, \theta_{1:m}, \alpha) = \frac{\pi(d_{1:m}, \theta_{1:m}, \alpha, \beta)}{\pi(d_{1:m}, \theta_{1:m}, \alpha)} \propto \pi(d_{1:m}, \theta_{1:m}, \alpha, \beta) \quad \dots\dots\dots (64)$$

$$\propto \prod_{i=1}^m \pi(d_i | \theta_i) \prod_{i=1}^m \pi(\theta_i | \alpha, \beta) \pi(\alpha) \pi(\beta) \quad \dots\dots\dots (65)$$

$$\propto \prod_{i=1}^m \pi(\theta_i | \alpha, \beta) \pi(\beta) \quad \dots\dots\dots (66)$$

となり、モデルの設定から、 $\pi(\theta_i | \alpha, \beta)$  及び  $\pi(\beta)$  は Gamma 分布であるので、

$$\propto \prod_{i=1}^m \left\{ \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\beta\theta_i) \right\} \frac{b_\beta^a}{\Gamma(a)} \beta^{a-1} \exp(-b_\beta \beta) \quad \dots\dots\dots (67)$$

$$\propto \beta^{m\alpha} \exp\left(-\beta \sum_{i=1}^m \theta_i\right) \beta^{a-1} \exp(-b_\beta \beta) \quad \dots\dots\dots (68)$$

$$\propto \beta^{(m\alpha+a)-1} \exp\left\{-\beta \left(\sum_{i=1}^m \theta_i + b_\beta\right)\right\} \quad \dots\dots\dots (69)$$

となる。したがって、 $\beta$  の完全条件付分布は、

$$\beta | d_{1:m}, \theta_{1:m}, \alpha \sim \text{Gamma}\left(m\alpha + a, \sum_{i=1}^m \theta_i + b_\beta\right) \quad \dots\dots\dots (70)$$

である。

まとめると、ギブス・サンプラーのための完全条件付分布は、次のようになる。

$$i) \theta_i | d_{1:m}, \theta_{-i}, \alpha, \beta \stackrel{\text{ind}}{\sim} \text{Gamma}(d_i + \alpha, e_i + \beta) \quad \dots\dots\dots (71)$$

$$\text{ii) } \pi(\alpha | d_{1:m}, \theta_{1:m}, \beta) \propto \frac{\beta^{m\alpha}}{\Gamma^m(\alpha)} \prod_{i=1}^m (\theta_i^{\alpha-1}) \exp(-b_\alpha \alpha) \quad \dots\dots\dots (72)$$

$$\text{iii) } \beta | d_{1:m}, \theta_{1:m}, \alpha \sim \text{Gamma} \left( m\alpha + a, \sum_{i=1}^m \theta_i + b_\beta \right) \quad \dots\dots\dots (73)$$

i), iii) は、直接的にサンプルを抽出するが、ii) は標準的な分布の形ではないので、**M-H**アルゴリズムによって抽出を行う。以上から、**Poisson-Gamma**モデルの**MCMC**法のアルゴリズムは表 1 のようにまとめられる<sup>25</sup>。

---

<sup>25</sup> ここで示したものは、**MCMC**法の基本的な手法に従ったアルゴリズムであり、より効率的なアルゴリズムは考えられうる。

表1 Poisson-Gamma モデルの MCMC 法

- (1) 初期値 ( $\alpha^{(0)}$ ,  $\beta^{(0)}$ ), M-H アルゴリズムの尺度パラメータ ( $\sigma^2$ ) を決める.  
 (2)  $j=1, 2, \dots$  について, 次を繰り返す (ギブス・サンプラー).

- (a)  $\theta_i^{(j)}$  ( $i=1, 2, \dots, m$ ) を発生させる.

$$\theta_i^{(j)} | d_{1:m}, \theta_{-i}, \alpha^{(j-1)}, \beta^{(j-1)} \stackrel{\text{ind}}{\sim} \text{Gamma}(d_i + \alpha^{(j-1)}, e_i + \beta^{(j-1)})$$

- (b)  $\beta^{(j)}$  を発生させる.

$$\beta^{(j)} | d_{1:m}, \theta_{1:m}^{(j)}, \alpha^{(j-1)} \sim \text{Gamma}\left(m\alpha^{(j-1)} + a, \sum_{i=1}^m \theta_i^{(j)} + b_\beta\right)$$

- (c)  $\alpha^{(j)}$  を発生させる (ランダムウォーク連鎖による M-H アルゴリズム).

- ① 新しい候補点  $\alpha^*$  を発生させる.

$$\alpha^* \sim \text{Normal}(\alpha^{(j-1)}, \sigma^2)$$

- ② 次の受理確率を計算する.

$$P(\alpha^{(j-1)}, \alpha^* | \theta_{1:m}^{(j)}, \beta^{(j)}) = \min[R, 1]$$

ここで,  $R = \exp\{\pi(\alpha^*) - \pi(\alpha^{(j-1)})\}$  であり, また

$$\pi(\alpha^*) = m\alpha^* \log \beta^{(j)} + (\alpha^* - 1) \sum_{i=1}^m \log \theta_i^{(j)} - b_\alpha \alpha^* - m \log \Gamma(\alpha^*) \text{ である.}$$

- ③  $u \sim \text{Uniform}(0, 1)$  を発生させ,  $\alpha^{(j)}$  を次のように決定する.

$$\alpha^{(j)} = \begin{cases} \alpha^* & \text{if } u \leq P(\alpha^{(j-1)}, \alpha^* | \theta_{1:m}^{(j)}, \beta^{(j)}) \\ \alpha^{(j-1)} & \text{else} \end{cases}$$

- (3) 十分大きな  $M$  に対して  $\theta_i^{(j)}$  ( $i=1, 2, \dots, m$ ),  $\beta^{(j)}$ ,  $\alpha^{(j)}$  ( $j=M, M+1, \dots$ ) を記録する.

MCMC法の実行に関して, ギブス・サンプラーの順序はサンプリングに影響しないため変更でき, 初期値には, 経験ベイズ法での最尤推定値やモーメント推定値を指定することができる. あるいは, 事後分布を近似するラプラス法などでの最適化アルゴリズムを使用して事後分布のモードと分散を指定する (Albert, 2009). M-Hアルゴリズムでは, 数値計算のアンダーフローを生じさせないために,  $\alpha$  の完全条件付分布における密度関数は対数密度関数  $\pi(\alpha^*)$  で記述している<sup>26</sup>. それにともない  $R$  は指数を

<sup>26</sup> 確率の積の計算では, 解が小さくなりすぎて桁が右にあふれてしまうことがある. この現象をアンダ

とっている。尺度パラメータは、受理率から調整する<sup>27</sup>。受理率は、提案密度を正規分布としたランダムウォーク連鎖では、25～45%が良いとされ、複数のパラメータを含む場合には25%が最良の選択とされる (Albert, 2009)。

## 7. 計算例：兵庫県の市区町村を対象とした相対リスクの推定

自殺の相対リスク推定の計算例として、2003-07年の兵庫県における男性の自殺発生数 (厚生労働省『人口動態統計』; 自殺予防総合対策センター) の相対リスクを推定することを検討してみることにする。兵庫県下47市区町村の相対リスクの最尤推定値と期待死亡数 (対数値) の関係は図1に示した。相対リスクの最尤推定量における一般的な傾向として、自殺発生数が少ない地域では散らばりが大きいことがわかる。

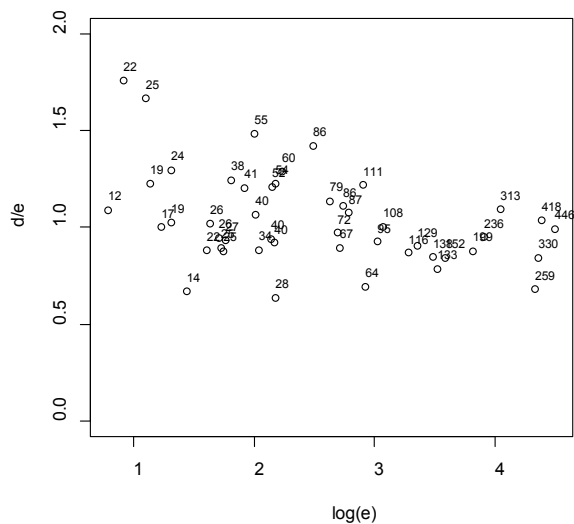


図1 2003-07年の兵庫県47市区町村における男性の自殺の相対リスクの最尤推定値

注：横軸は期待死亡数の対数，縦軸は相対リスク，数字は自殺発生数を示す。

次に、ベイズ推定による相対リスクの縮約推定量を求めてみることにする。計算では、R (R Core Team, 2013; ver. 3.0.1) を使用した<sup>28</sup>。経験ベイズ法の最尤推定値は、

---

ーフローと呼ぶ。この問題を避けるため、確率の積の計算は対数をとることで和の計算にすることが一般的に行われる。

<sup>27</sup> 受理率は、反復を通しての受理確率の平均である。MCMC サンプル全体における受理されたサンプルの割合で表される。

<sup>28</sup> R による最尤推定法や MCMC 法の設計においては、Rizzo (2008) を参考とした。

対数尤度関数を指定してoptim関数による数値計算によって求めた<sup>29</sup>. 地域全体を兵庫県として、 $\alpha$ と $\beta$ の初期値はともに5とした。

階層ベイズ法は、表1のアルゴリズムに従い、RでMCMC法を実行した<sup>30</sup>. MCMCアルゴリズムの収束診断には、Rのcoda (Plummer, Best, Cowles, & Vines, 2006) パッケージに含まれる関数を使用した. 連鎖数を3つとしてそれぞれ10万個発生させ、最初の2000個を初期値に依存する期間として捨て、さらにバッチサイズ10で間引き、それぞれの連鎖で10000個の標本を記録した. 初期値は、 $\alpha^{(0)}$ と $\beta^{(0)}$ 共に10, 75, 150 (3つの連鎖) とし、M-Hアルゴリズムの尺度パラメータ $\sigma^2 = 3.24$ とした.

MCMC法によるサンプリングの結果として、トレースプロットと事後確率密度関数を図2, 自己相関関数のプロットを図3に示した. M-Hアルゴリズムの受率率は、24.0%であった.

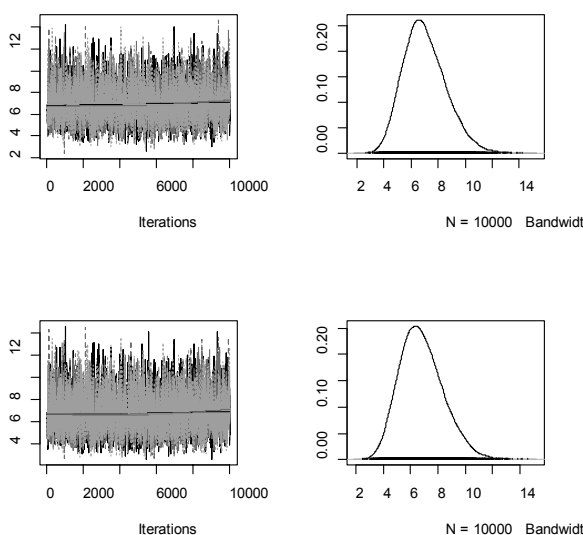


図2 トレースプロットと事後密度関数のプロット

注：上段 $\alpha$ ，下段 $\beta$ を示している。

<sup>29</sup> 経験ベイズ法（最尤推定）のRコードは、付録3で示した。

<sup>30</sup> 階層ベイズ法のRコードは、付録4で示した。

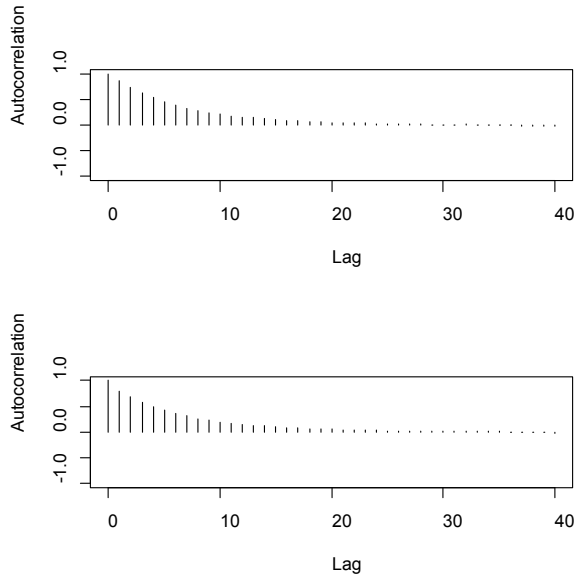


図3 自己相関関数のプロット

注：上段  $\alpha$ ，下段  $\beta$ ，最初の連鎖における自己相関関数を示している。

収束診断を行うと、トレースプロットはランダムなノイズに近く、自己相関はラグの関数として減衰傾向を示している。また、Geweke の  $p$  値はいずれも 0.05 より大きく、Gelman & Rubin の  $\hat{R}$  指数はともに 1 であり、サンプリングは収束していると判断した。

表 2 には、得られた MCMC サンプルから計算された  $\alpha$  と  $\beta$  の事後分布の平均、標準偏差、95%信用区間の下限と上限の値、平均の標準誤差を示している。

表 2 階層ベイズ法での  $\alpha$ ， $\beta$  の推定結果

	平均	標準偏差	95%下限	95%上限	Naive SE	Time-series SE
$\alpha$	7.012	1.572	4.334	10.445	0.009	0.032
$\beta$	6.812	1.603	4.097	10.299	0.009	0.031

階層ベイズ法では平均では  $\alpha=7.012$ ， $\beta=6.812$  であり、この値での相対リスクは平均 1.029、分散 0.151 である。経験ベイズ法での推定結果は、 $\alpha=142.448$ ， $\beta=148.560$  であり、この値での相対リスクは平均 0.959、分散 0.006 である。表 3 として、最尤推定値、経験ベイズ推定値、階層ベイズ推定値を示し、図 4 には階層ベイズ推定値と最尤推定値を示した。階層ベイズ推定値は、人口が多い地域では最尤推定値とほぼ異ならないが、人口が少ない市区町村では中央付近に縮約されていることが明確である。

表 3 相対リスクの最尤推定値，経験ベイズ推定値，階層ベイズ推定値

市区町村名	期待 死亡数	自殺 発生数	最尤 推定値	経験ベイズ 推定値	階層ベイズ 推定値
神河町	2.2	2.4	1.091	0.961	1.049
福崎町	3.4	3.4	1.000	0.960	1.022
相生市	5.7	5	0.877	0.956	0.961
宍粟市	7.4	11	1.486	0.984	1.271
南あわじ市	9.3	12	1.290	0.978	1.183
芦屋市	15	13.4	0.893	0.953	0.935
三田市	18.5	12.8	0.692	0.929	0.782
神戸市須磨区	28.6	25.8	0.902	0.950	0.927
加古川市	45.3	39.8	0.879	0.940	0.898
神戸市垂水区・西区	78.4	66	0.842	0.918	0.857

注：市区町村 10 地域のみ抜き出し示している。

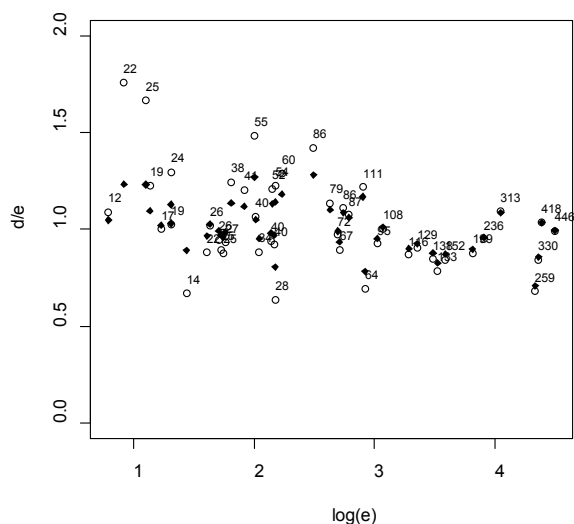


図 4 階層ベイズ推定値（黒点）と最尤推定値（白点）

注：横軸は期待死亡数の対数，縦軸は相対リスク，数字は自殺発生数を示す。

## 8. おわりに

本稿では，自殺の実態を把握する計量的指標として相対リスクを取り上げ，その推定方法として，最尤推定量，経験ベイズ推定量，階層ベイズ推定量について，理論的にそして実証的に検討を行った。

自殺発生数にポアソン分布を仮定した相対リスクの最尤推定量は，理論的には有効推定量であるが，人口の少ない地域において標本誤差が大きくなるという性質がある。この実証的な面での問題を回避するために，事前分布にガンマ分布を仮定した相対リ



スクの基本となる Poisson-Gamma モデルを扱った。事前分布を仮定するベイズ推定の枠組みが縮約推定量を構成することを説明した。つまり、ベイズ推定量は、地域全体の推定の力を借りることで、各地域の推定をより適切にするといえる。

ベイズ推定においては、事前分布の設定方法で 2 種類の推定法を検討した。すなわち、経験ベイズ法と階層ベイズ法である。前者は、データを利用して事前分布の定数パラメータを推定する。これに対して、後者は、事前分布のパラメータについても定数ではなく確率変数として、さらなる事前分布（超事前分布）を想定する。そのため、階層ベイズ法では、多次元の積分が必要になり、解析的に求まらない事後分布からのサンプリングを可能にする MCMC 法によるサンプリングが必要となった。Poisson-Gamma モデルの事後分布を求めるため、ギブス・サンプラーによるサンプルを抽出するための完全条件付分布を導き、ギブス・サンプラーと M-H アルゴリズムを用いたアルゴリズムを示した。

自殺の相対リスクにおける各推定値の理論的な面での特徴は次のように要約できる。まず、最尤推定値は、有効推定量であり推定量が持つべき望ましい性質を備えている。しかしその一方で、データへの依存が大きく小地域での推定精度が低くなるため、実態をあらわす正確な指標ではない。次に、階層ベイズ推定値は、縮約推定値としての性質により、小地域でも精度の高い指標である。さらに、事後分布を構成することで信用区間の構成が容易であり、指標の安定性を把握することを可能にする。ただし、事前分布の選択においては検討を要する。したがって、研究者のモデルの構成が適切であれば、小地域でも精度が高く安定した情報をもたらす。そして、経験ベイズ推定値は、階層ベイズ推定値と最尤推定値の中間の特長を持つと言える。つまり、ベイズ推定値の特長として小地域の精度は高いが、一方では事前情報を完全には活かしておらずデータに依存している近似的な手法である。

本稿では、相対リスクの推定における最も基本的な Poisson-Gamma モデルを取り上げたが、Poisson-Gamma モデルの 1 つの欠点は、空間や時間での関連を加味することができない点である。これに対処するため、事前分布にガンマ分布ではなく対数正規分布を仮定する Log-normal モデル、条件付自己回帰（conditional autoregressive；CAR）モデルなどの発展的なモデルが存在している（Lawson, 2013; Rao, 2003）。これらのモデルも今回提示したようなベイズ推定法の構成で行われている。これらに関しては今後検討していきたい。

#### 【謝 辞】

本研究は、関西大学・社会的信頼システム創生プロジェクト（文部科学省私立大学戦略的研究基盤形成支援事業：平成 22 年度～平成 26 年度）の一部として行われたものである。本論文を作成するにあたり、与謝野有紀先生（関西大学社会学部）、清水和秋先生（関西大学社会学部）のご指導、ご助言を頂きました。深く感謝いたします。

#### 【引用文献】

- Albert, J., 2009, *Bayesian computation with R*, Philadelphia: Springer Science+Business Media. (=2012, 石田基広・石田和枝訳『Rで学ぶベイズ統計学入門』丸善出版.)
- Clayton, D. & Kaldor, J., 1987, “Empirical bayes estimates of age-standardized relative risks for use in disease mapping,” *Biometrics*, 43: 671-681.
- Datta, G., Ghosh, M. & Waller, L. A., 2000, “Hierarchical and empirical Bayes methods for environmental risk assessment”, Sen P. K. & Rao C. R. eds., *Handbook of Statistics, Bioenvironmental and Public Health Statistics 18 edition*, Amsterdam: Elsevier, 223-245.
- Efron, B. & Morris, C., 1973, “Stein’s estimation rule and its competitors- an empirical Bayes approach,” *Journal of American Statistical Association*, 68(341): 117-130.
- Gelman, A. & Rubin, D., 1992, “Inference from iterative simulation using multiple sequences (with discussion),” *Statistical Science*, 7(4): 457-511.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B., 2013, *Bayesian Data Analysis, Third Edition*, Broken Sound Parkway NW: CRC Press.
- George, E. I., Makov, U. E., & Smith, A. F. M., 1993, “Conjugate likelihood distributions,” *Scandinavian Journal of Statistics*, 20(2): 147-156.
- Geweke, J., 1992, “Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion),” Bernardo, J., Berger, J., Dawid, A., and Smith, A. eds., *Bayesian Statistics 4*, Oxford: Oxford University Press, 169-193.
- Gordis, L., 2009, *Epidemiology, fourth edition*, Philadelphia: Saunders. (=2010, 木原正博・木原雅子・加治正行訳『疫学——医学的研究と実践のサイエンス』メディカル・サイエンス・インターナショナル.)
- 伊庭幸人, 2005, 「マルコフ連鎖モンテカルロ法の基礎」伊庭幸人・種村正美・大森裕浩・和合肇・佐藤整尚・高橋明彦『計算統計Ⅱ マルコフ連鎖モンテカルロ法とその周辺』岩波書店, 1-106.
- Jacquier, E. & Polson, N., 2011, “Bayesian Methods in Finance,” Geweke, J., Koop, G., & van Dijk, H. eds., *The Oxford Handbook of Bayesian Econometrics, First Edition*, Oxford: Oxford University Press. (=2013, 島田淳二訳「ベイズ統計のファイナンスへの応用」照井伸彦監訳『ベイズ計量経済学ハンドブック』朝倉書店, 448-528.)
- James, W. & Stein, C., 1961, “Estimation with quadratic loss,” *Proc. Fourth*

- Berkeley Symp. Math. Statist. Prob.*, 1: 361–379.
- 久保拓弥, 2012, 『データ解析のための統計モデリング入門——一般化階層モデル・階層ベイズモデル・MCMC』岩波書店.
- 久保川達也, 2004, 「スタインのパラドックスと縮約推定の世界」甘利俊一・竹内啓・竹村彰通・伊庭幸人編『統計科学のフロンティア 3 モデル選択——予測・検定・推定の交差点』岩波書店, 139-197.
- Lancaster, T., 2004, *An Introduction to Modern Bayesian Econometrics*, Oxford: Blackwell Publishing Ltd.. (=2011, 小暮厚之・梶田幸作監訳『ランカスター ベイジアン計量経済学』朝倉書店.)
- Lawson, A. B., 2013, *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology, Second Edition*, London: Chapman & Hall/CRC.
- Link, W. A. & Hahn, D. C., 1996, “Empirical Bayes estimation of proportions with application to cowbird parasitism rates,” *Ecology*, 77: 2528-37.
- Maiti, T., 2005, “Bayesian Aspects of Small Area Estimation,” Day, D. K. & Rao, C. R. eds., *Handbook of statistics volume 25 Bayesian Thinking: Modeling and Computation*, Amsterdam: Elsevier, 965-982. (=2011, 土居正明訳「ベイズ統計を用いた小地域推定」繁榘算男・岸野洋久・大森裕浩監訳『ベイズ統計分析ハンドブック』朝倉書店, 1003-1023.)
- Marshall, R. J., 1991, “Mapping disease and mortality rates using empirical bayes estimators,” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 40(2): 283-294.
- 松原望, 2010, 『ベイズ統計学概説——フィッシャーからベイズへ』培風館.
- Morris, C. N., 1983, “Parametric empirical bayes inference: theory and applications,” *Journal of the American Statistical Association*, 78: 47-55.
- Natarajan, K., Ghosh, M., & Maiti, T., 1998, “Hierarchical bayes quality measurement plan,” *Communications in Statistics – Simulation and Computation*, 27(1): 199-214.
- 大森裕浩, 2001, 「マルコフ連鎖モンテカルロ法の最近の展開」『日本統計学会誌』31(3): 305-344.
- Plummer, M., 2013, *JAGS (Just Another Gibbs Sampler)*, Version 3.4.0. <http://mcmc-jags.sourceforge.net/>.
- Plummer, M., Best, N., Cowles, K. & Vines, K., 2006, “CODA: Convergence Diagnosis and Output Analysis for MCMC,” *R News*, 6: 7-11.
- Qian, S. S., 2010, *Environmental and Ecological Statistics with R*, London: Chapman & Hall/CRC. (=2011, 大森浩二・井上幹生・畑 啓生監訳『環境科学と生態学のための R 統計』共立出版.)

- R Core Team, 2013, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.  
<http://www.R-project.org/>.
- Rao, J. N. K., 2003, *Small area estimation*, NJ: John Wiley & Sons.
- Rizzo, M., 2008, *Statistical Computing with R*, : CRC Press. (=2011, 石井一夫・村田真樹訳『Rによる計算機統計学』オーム社.)
- Robert, C. & Casella, G., 2010, *Introducing Monte Carlo Methods with R*, : Springer Science+Business Media. (=2012, 石田基広・石田和枝訳『Rによるモンテカルロ法入門』丸善出版.)
- Rothman, K. J., 2012, *Epidemiology: An Introduction, 2nd edition*, Oxford: Oxford University Press. (=2013, 矢野栄二・橋本英樹・大脇和浩監訳『ロスマンの疫学——科学的思考への誘い 第2版』篠原出版新社.)
- Singer, J. D. & Willett, J. B., 2003, *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford: Oxford University Press. (=2012, 菅原ますみ監訳『縦断データの分析 I ——変化についてのマルチレベルモデリング』朝倉書店.)
- Stan Development Team, 2014, *Stan: A C++ Library for Probability and Sampling*, Version 2.2. <http://mc-stan.org>.
- Stein, C., 1956, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," *Proc. Third Berkeley Symp. Math. Statist. Prob.*, 1: 197–206.
- 竹村彰通, 1991, 『現代数理統計学』創文社.
- 丹後俊郎・横山徹爾・高橋邦彦, 2007, 『空間疫学への招待—疾病地図と疾病集積性を中心として』朝倉書店.
- 和合肇編, 2005, 『ベイズ計量経済分析—マルコフ連鎖モンテカルロ法とその応用』東洋経済新報社.

## 【付録 1】 周辺尤度の導出

周辺尤度<sup>31</sup>は,

$$p(d_i; \alpha, \beta) = \int_{\theta_i} p(d_i | \theta_i) g(\theta_i; \alpha, \beta) d\theta_i \quad \dots\dots\dots (A.1)$$

であり, 潜在的な相対リスク  $\theta_i$  の定義域は  $0 \leq \theta_i \leq \infty$  であるため, 積分区間は  $\theta_i: 0 \rightarrow \infty$  となり, また, (14)式と(16)式を代入し,

$$= \int_0^\infty \frac{(e_i \theta_i)^{d_i} \exp(-e_i \theta_i)}{d_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\beta \theta_i) d\theta_i \quad \dots\dots\dots (A.2)$$

$$= \frac{e_i^{d_i} \beta^\alpha}{d_i! \Gamma(\alpha)} \int_0^\infty \theta_i^{d_i + \alpha - 1} \exp\{-(e_i + \beta)\theta_i\} d\theta_i \quad \dots\dots\dots (A.3)$$

となる. ここで, 積分部分をガンマ関数の形式に持ち込むため,  $(e_i + \beta)\theta_i = t$  とおき,  $\theta_i$  を  $t$  とする置換積分を行う. 積分区間は,  $\theta_i: 0 \rightarrow \infty$  から  $t: 0 \rightarrow \infty$  となり,  $(e_i + \beta)d\theta_i = dt$  から  $d\theta_i = (e_i + \beta)^{-1} dt$  となるので, (A.3)式は,

$$= \frac{e_i^{d_i} \beta^\alpha}{d_i! \Gamma(\alpha)} \int_0^\infty \left( \frac{t}{e_i + \beta} \right)^{d_i + \alpha - 1} \exp(-t) (e_i + \beta)^{-1} dt \quad \dots\dots\dots (A.4)$$

$$= \frac{1}{d_i! \Gamma(\alpha)} \frac{e_i^{d_i} \beta^\alpha}{(e_i + \beta)^{d_i + \alpha}} \int_0^\infty t^{d_i + \alpha - 1} \exp(-t) dt \quad \dots\dots\dots (A.5)$$

積分部分は, ガンマ関数として  $\Gamma(d_i + \alpha)$  となるため<sup>32</sup>,

$$= \frac{\Gamma(d_i + \alpha)}{d_i! \Gamma(\alpha)} \frac{e_i^{d_i} \beta^\alpha}{(e_i + \beta)^{d_i + \alpha}} \quad \dots\dots\dots (A.6)$$

$$= \frac{\Gamma(d_i + \alpha)}{d_i! \Gamma(\alpha)} \left( \frac{\beta}{e_i + \beta} \right)^\alpha \left( \frac{e_i}{e_i + \beta} \right)^{d_i} \quad \dots\dots\dots (A.7)$$

となる. ガンマ関数を階乗の形式になおして<sup>33</sup>,

$$= \frac{(d_i + \alpha - 1)!}{d_i! (\alpha - 1)!} \left( \frac{\beta}{e_i + \beta} \right)^\alpha \left( \frac{e_i}{e_i + \beta} \right)^{d_i} \quad \dots\dots\dots (A.8)$$

$$= \binom{d_i + \alpha - 1}{d_i} \left( \frac{\beta}{e_i + \beta} \right)^\alpha \left( \frac{e_i}{e_i + \beta} \right)^{d_i} \quad \dots\dots\dots (A.9)$$

となる. したがって, 周辺尤度は,  $d_i \stackrel{\text{iid}}{\sim} \text{Negative Binomial}(\alpha, \beta)$  である<sup>34</sup>.

<sup>31</sup> 周辺尤度は,  $p(y) = \int_{\theta} p(y, \theta) d\theta = \int_{\theta} p(y | \theta) p(\theta) d\theta$  で求められる.

<sup>32</sup> ガンマ関数は,  $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) dx$  である.

<sup>33</sup> ガンマ関数の性質として,  $\Gamma(\alpha) = (\alpha - 1)!$  である.

<sup>34</sup> 負の二項分布の確率関数は,  $p(x) = \binom{x+r-1}{x} p^\alpha q^x$ ,  $x = 0, 1, 2, \dots$ ,  $p + q = 1$  である.

【付録 2】 事後分布の導出

ベイズの定理より,

$$w(\theta_i | d_i; \alpha, \beta) = \frac{p(d_i | \theta_i) g(\theta_i; \alpha, \beta)}{\int_{\Theta_i} p(d_i | \theta_i) g(\theta_i; \alpha, \beta) d\theta_i} \quad \dots\dots\dots (B.1)$$

である. ここで, 分子は, (14)と(16)を代入して,

$$p(d_i | \theta_i) g(\theta_i; \alpha, \beta) = \frac{(e_i \theta_i)^{d_i} \exp(-e_i \theta_i)}{d_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\beta \theta_i) \quad \dots (B.2)$$

である. また, 分母は  $d_i$  の周辺尤度であり, (A.6)式を代入して,

$$\int_{\Theta_i} p(d_i | \theta_i) g(\theta_i; \alpha, \beta) d\theta_i = \frac{\Gamma(d_i + \alpha)}{d_i! \Gamma(\alpha)} \frac{e_i^{d_i} \beta^\alpha}{(e_i + \beta)^{d_i + \alpha}} \quad \dots\dots\dots (B.3)$$

である. したがって, (B.1)式は,

$$w(\theta_i | d_i; \alpha, \beta) = \frac{\frac{(e_i \theta_i)^{d_i} \exp(-e_i \theta_i)}{d_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\beta \theta_i)}{\frac{\Gamma(d_i + \alpha)}{d_i! \Gamma(\alpha)} \frac{e_i^{d_i} \beta^\alpha}{(e_i + \beta)^{d_i + \alpha}}} \quad \dots\dots\dots (B.4)$$

となり, 約分して,

$$w(\theta_i | d_i; \alpha, \beta) = \frac{(\beta + e_i)^{\alpha + d_i}}{\Gamma(\alpha + d_i)} \theta_i^{\alpha + d_i - 1} \exp\{-(\beta + e_i)\theta_i\} \quad \dots\dots\dots (B.5)$$

となる. これは,  $\theta_i | d_i \stackrel{\text{ind}}{\sim} \text{Gamma}(\alpha + d_i, \beta + e_i)$  である.

### 【付録3】 経験ベイズ法（最尤推定法）のためのRコード

```
# Poisson-Gamma モデルの最尤推定
# データの読み込み
data <- read.csv("hyougo.csv",header=T)
d <- data$d
e <- data$e
m <- length(d)

# 目的関数（対数尤度関数）の設定，母数は theta=c(alpha,beta)
LL <- function(theta,d,e,m)
{
  alpha <- theta[1]
  beta <- theta[2]
  stand <- numeric(m)
  for(i in 1:m)
  {
    n <- d[i] + alpha - 1
    k <- d[i]
    logch <- lchoose(n,k) # 二項係数の対数
    stand[i] <- logch + alpha*log(beta) - (d[i]+alpha)*log(e[i]+beta) + d[i]*log(e[i])
  }
  loglik <- sum(stand)
  -loglik
}

# 最適化（Nelder-Mead 法）
op <- optim(c(5,5), LL, d=d, e=e, m=m, hessian=TRUE)
```

#### 【付録 4】 階層ベイズ法の MCMC 法のための R コード

```
data <- read.csv("hyougo.csv",header=T) # データの読み込み
library(coda) # MCMC 収束診断に使用するライブラリーの読み込み
## MCMC
# データと事前分布パラメータの設定
d <- data$d # 自殺死亡数のベクトル
e <- data$e # 期待死亡数のベクトル
m <- nrow(data) # 地域の数
a1 <- 1 # alpha の事前分布 Gamma(a1,b1)の超パラメータ a1
b1 <- 1 # alpha の事前分布 Gamma(a1,b1)の超パラメータ b1
a2 <- 0.1 # beta の事前分布 Gamma(a2,b1)の超パラメータ a2
b2 <- 1 # beta の事前分布 Gamma(a2,b2)の超パラメータ b2

# alpha の完全条件付分布の対数密度関数
fa <- function(alpha,beta,theta,a,b)
{
  slthe <- sum(log(theta))
  logf <- (m*alpha)*log(beta) + (alpha-1)*slthe - b*alpha - m*loggamma(alpha)
  return(logf)
}

# MCMC の設定
chain <- 3 # 連鎖の数
alphst <- c(10,75,150) # alpha の初期値 (連鎖の数だけ準備する)
betast <- c(10,75,150) # beta の初期値 (連鎖の数だけ準備する)
# burn <- 10000 # バーンイン期間
# thin <- 10 # サンプリング間隔 (間引き)
Niter <- 102000 # MCMC の繰り返し数 (サンプリング数)
reall <- array(0, dim=c((Niter*chain), (m+3))) # 結果の格納行列を準備

# MCMC の実行
for (t in 1:chain) # ここで t は, MCMC サンプル連鎖の数 t(1,2,...,chain)である
{
  # MCMC の初期化
  alpha <- numeric(Niter) # alpha のサンプル格納ベクトルを準備
  beta <- numeric(Niter) # beta のサンプル格納ベクトルを準備
  theta <- array(0, dim=c(Niter, m)) # theta のサンプル格納行列を準備
  alpha[1] <- alphst[t] # alpha 初期値
  beta[1] <- betast[t] # beta の初期値
  #theta[1,] <- rep(0,m) # theta の初期値(今回の連鎖順番では必要なし)
  sigma <- 1.9 # ランダムウォークの尺度母数 (正規分布の標準偏差, 採択率で調整する)
  u <- runif(Niter) # 受理・棄却用の一様分布の生成
  k <- numeric(Niter) # 受理数の格納ベクトルを準備

  # ギブス・サンプラーを実行する
  for (i in 2:Niter) # ここで i は, MCMC サンプル数 i(2,3,...,Niter)である
  {
    # theta を生成する
    for (j in 1:m) # ここで j は, 地域 j(1,2,...,m)である
    {
      theta[i,j] <- rgamma(1, shape=d[j] + alpha[i-1], rate=e[j] + beta[i-1])
    }

    # beta を生成する
    sumthe <- sum(theta[i,])
    beta[i] <- rgamma(1, shape=m*alpha[i-1] + a1, rate=sumthe + b2)
  }
}
```



```

# alpha を生成する (正規ランダムウォークのメトロポリス・アルゴリズム)
y      <- rnorm(1, alpha[i-1], sigma) # 提案値の生成
logr   <- fa(y, beta[i], theta[i,], a2, b2) - fa(alpha[i-1], beta[i], theta[i,], a2, b2)
if(u[i] <= exp(logr))
{
  alpha[i] <- y
  k[i]     <- 1
}
else
{
  alpha[i] <- alpha[i-1]
}
}

# MCMC サンプルング結果の記録
result <- cbind(alpha=alpha, beta=beta, theta=theta, k=k)
reall[[(t-1)*Niter+1):(t*Niter),] <- result
}
# MCMC 終了

```

## Application of Bayesian Method to Estimating Relative Risks of Suicide in Japan

- Making Shrinkage Estimators of Standard Mortality Rate Using Empirical Bayesian Models and Hierarchical Bayesian Models -

Hiroaki KONDA

### 【Abstract】

Japanese suicide rate got drastically higher in 1998 and it is keeping high since then. And prevention of suicide is getting a most important political issue in contemporary Japan. In terms of effective preventions of suicide, assessing suicide risks in small areas have been regarded as an essential task, and standardized mortality rates (SMR) is generally used to showing pictures of conditions of suicide in small areas. However, SMR has disadvantages of showing excess variability in calculating it for smaller population areas such as rural districts. Therefore, getting reliable estimators of SMR is one of the most important tasks for suicide researchers and policy makers until now. In this study, we focused on maximum likelihood estimators, empirical Bayesian estimators, and hierarchical Bayesian estimators of SMR as indicators of relative suicide risks. First, we applied basic Poisson-Gamma model for areas in Hyogo prefecture in Japan and discussed that the Poisson likelihood model with Gamma prior distribution constructed the shrinkage estimators. Second, we showed how to construct hierarchical Bayesian estimator with Markov chain Monte Carlo methods. Finally, we provided a comprehensive discussion of these three estimators' features.

Keywords and Phrases: suicide, hierarchical Bayesian models, MCMC, small area estimation