

市区町村別にみた死因別死亡率のベイズ推定について

ハセガワ イサオ
長谷川 功*1

目的 市区町村別生命表のような小規模地域における死亡状況を表す指標を算出する際に死因分析を行うことを目標として、小地域の不安定性を解消し、かつ、全死因については従来の推定方法との整合性を持つような死因別死亡率の推定方法を導入し、算出結果を検討する。

方法 従来の市区町村別生命表において中央死亡率をベイズ推定する際に用いるモデル（尤度関数を二項分布，事前分布をベータ分布）を多変数に拡張したモデルとして知られている多項分布とディリクレ分布との組を死因別中央死亡率のベイズ推定モデルとすることを提案する。さらに，導入したモデルにより算出される死因別中央死亡率（ベイズ推定値）を用いて死因を除去した場合の平均余命の延びと死因別死亡確率を試算して，モデルの有用性を検討した。

結果 試算した2種類の指標について，人口規模の小さな市区町村に対しては偶然変動の影響を抑えた結果が得られ，その一方で，人口規模の大きな市区町村に対しては，ベイズ推定を用いない方法とほぼ等しい結果が得られた。

結論 本稿で作成したベイズ推定値は小地域の死亡状況の分析手法として今後検討を重ねて活用されることが期待される。また，生命表以外の死因分析についても本稿と同様な方法で指標を作成できるかどうか今後研究する価値がある。

キーワード 市区町村別生命表，ベイズ推定，死因分析，平均余命の延び，死因別死亡確率

はじめに

厚生労働省大臣官房統計情報部では，死亡状況を分析する指標のひとつとして，各種の生命表を作成・公表している。生命表は作成の基礎となる地域の年齢構成によらない死亡指標であり，わが国の死亡状況を表すものとして，毎年作成される簡易生命表と，5年ごとに作成される完全生命表がある。一方，地域の特性にあう行政を行うための政策立案の資料を提供するために，都道府県別生命表，市区町村別生命表を5年ごとに作成しており，本号別稿「生命表と年齢調整死亡率の関係について」や「都道府県

別生命表による平均寿命の地域差分析」にみられるような地域間の死亡状況の比較の際の基礎資料として活用されている。

地域別の指標の中でも，特に小規模な地域に対する指標となる市区町村別生命表では，本号別稿「最近のベイズ推定研究の小地域の人口動態指標推定への応用の研究」で考察しているような指標の不安定性を有している。この不安定性を解消するため，従来の市区町村別生命表では，性・年齢階級別の中央死亡率をベイズ推定により求めている。ここで，中央死亡率とは，作成基礎期間の中央時点の人口（以下，中央人口）に対する期間中の死亡数の割合として求め

*1 厚生労働省大臣官房統計情報部企画課審査解析室総合解析係

（現 = 内閣府政策統括官（経済財政分析担当）付参事官（総括担当）付公共経済班）

られる値である。現実に観察される事象は人口および死亡数であるので、ベイズ推定のモデルとしては、中央死亡率を確率変数とみなして、中央死亡率に従う反復試行を中央人口だけ繰り返した時に成功した試行回数が死亡数に等しいとするモデルを適用している。つまり、尤度関数として二項分布を用いてベイズ推定を行うことにする。さらに、確率変数とみなした中央死亡率の事前分布として、比率を表す確率変数の事前分布として有用な分布のひとつであるベータ分布を採用しており、二項分布とベータ分布の共役性から事後分布もベータ分布となる事がわかる。上述のベイズ推定モデル（二項分布・ベータ分布モデル）を用いて、生命表を作成する地域ごとに性・年齢階級別の中央死亡率（ベイズ推定値）を求めることによって、出現数の少なさに起因する偶然性の影響を減少させることが可能となり、市区町村別生命表の諸生命表関数は安定性の高い数値となっている。

一方で、生命表作成の基礎数値である人口動態統計では、死亡数は死因別に集計されており、簡易生命表や都道府県別生命表では、平均余命などの生命表関数を分析する手法のひとつとして、特定死因を除去した場合の平均余命の延びや死因別死亡確率などの指標を死因別死亡数を用いて算出している。小地域の指標に対しても同様に死因別データを活用することが期待されており、統計情報部が作成している「人口動態保健所・市区町村別統計」においては主要死因別の標準化死亡比が算出・公表されているが¹⁾、市区町村別生命表において、死因はこれまで考慮されていなかった。

本稿では、市区町村別の死亡率を死因まで考慮して推定することを目標とする。このとき死因別のデータを用いて死亡率を算出することになるが、死因を忘れてしまえば従来の推定で用いた情報と一致することから、全死因については従来の市区町村別生命表における推定方法との整合性を持ち、その上で、小地域の不安定性を解消するような推定方法が望まれる。そこで、市区町村別生命表で用いている二項分布・ベータ分布モデルを多変数分布に拡張したものと

て知られている多項分布とディリクレ分布との共役な分布の組（多項分布・ディリクレ分布モデル）をベイズ推定のモデルとすることを提案する。この方法により死因別中央死亡率（ベイズ推定値）を求めることが可能となり、さらに、死因に対する加法性を満たす結果が得られた。本稿ではまず第 節で、多項分布・ディリクレ分布モデルによる死因別中央死亡率（ベイズ推定値）の推定方法を導入し、第 節でその結果得られる死因を除去した場合の平均余命の延びや死因別死亡確率の計算方法について説明する。そして第 節で計算結果を示し、本稿で作成した指標の性質を検討する。

多項分布・ディリクレ分布モデル

以下、死因別中央死亡率（ベイズ推定値）を求めるために必要な理論的背景について解説する。ある集団の中央人口および死因別死亡数を、

人口： p ,

第 i 死因死亡数： $d_i, i=1 \dots k$,

死亡総数： $d_0 = \sum_{i=1}^k d_i$

とおく。ここで、死因については重複しないように分類しておくこととし、死因の分類数 k は固定しておく。

第 i 死因による死因別中央死亡率を x_i としたとき、ベイズ推定においては死亡率 x_i をある確率分布にしたがう確率変数として捉える。この確率変数は、観測によって人口 p および死亡数 $d_1 \dots d_k$ を得る以前では、ある分布（事前分布）に従っているとみなし、観測によって新たな情報が得られると、その情報によって確率変数 x_i が従う分布が更新される（事後分布）。死因別中央死亡率を観測によって得たデータから推定するときは推定値として事後分布の期待値で代表させることにして、その期待値を死因別中央死亡率（ベイズ推定値）とする。

多項分布・ディリクレ分布モデルにおいては、ベイズ推定における尤度関数は、

$$\Pr(d | x) = \frac{d_0!}{d_1!d_2!\dots d_k!(p-d_0)!} x_1^{d_1}x_2^{d_2}\dots x_k^{d_k}(1-x_0)^{p-d_0}$$

と表されると仮定する。ここで $d = (d_1, d_2, \dots, d_k)$ かつ $x = (x_1, x_2, \dots, x_k)$ とする。上述の尤度関数においては、人口 p は死亡数と同様にパラメータの1つではあるが、死亡数に比べ変動が小さいと考え定数として扱う。さらに、尤度関数が多項分布で表されると仮定しているため、共役分布であるディリクレ分布を事前分布として用いれば事後分布もディリクレ分布となることがわかり、事後分布の期待値、つまり死因別中央死亡率（ベイズ推定値）を算出することが容易になる。記号を簡単にするために $(k+1)$ 次元ディリクレ分布の密度関数を

$$\text{Dir}(x, y | \alpha_1, \dots, \alpha_k, \alpha_{k+1}) = C \cdot x_1^{\alpha_1-1} x_2^{\alpha_2-1} \dots x_k^{\alpha_k-1} y^{\alpha_{k+1}-1}$$

で表す。ここで、 k 個の独立変数と従属変数、およびパラメータは以下の条件式

$$\begin{aligned} x_i > 0, i = 1, \dots, k, \\ x_1 + \dots + x_k < 1, \\ y = 1 - (x_1 + \dots + x_k), \\ \alpha_i > 0, i = 1, \dots, k \\ \alpha_{k+1} > 0 \end{aligned}$$

を満たし、係数 C は正規化定数であり、ガンマ関数 $\Gamma(\cdot)$ を用いて、

$$C = \frac{\Gamma(\alpha_1 + \alpha_2 + \dots + \alpha_k + \alpha_{k+1})}{\Gamma(\alpha_1)\Gamma(\alpha_2)\dots\Gamma(\alpha_k)\Gamma(\alpha_{k+1})}$$

と表される。上記の定義式から $k=1$ のときは、多項分布が二項分布に、ディリクレ分布の密度関数がベータ分布の密度関数に一致していることがわかる。多項分布・ディリクレ分布モデルにおける事後分布の密度関数 $f(x | d)$ は、ベイズの定理から、

$$\begin{aligned} f(x | d) &= \frac{\Pr(d | x) \cdot \text{Dir}(x, y | \alpha_1, \dots, \alpha_k, \alpha_{k+1})}{\Pr(d | x) \cdot \text{Dir}(x, y | \alpha_1, \dots, \alpha_k, \alpha_{k+1}) dx_1 \dots dx_k} \\ &= \text{Dir}(x, y | \alpha_1 + d_1, \dots, \alpha_k + d_k, \alpha_{k+1} + p - d_0) \end{aligned}$$

となることがわかる。

一般のディリクレ分布 $\text{Dir}(x, y | \alpha_1, \dots, \alpha_k, \alpha_{k+1})$ に対して、変数 x_i の期待値および分散が、

$$E(x_i) = \frac{\alpha_i}{\alpha_0 + 1}$$

$$\begin{aligned} V(x_i) &= \frac{\alpha_i(\alpha_0 + 1 - \alpha_i)}{(\alpha_0 + 1)^2(\alpha_0 + 2)} \\ &= \frac{E(x_i)(1 - E(x_i))}{\alpha_0 + 2} \end{aligned}$$

となることが知られているので、多項分布・ディリクレ分布モデルにおける事後分布の期待値および分散が事前分布のパラメータ α_i と観測値 d_i, p から容易に求められることがわかる。

死因別中央死亡率をベイズ推定するために多項分布・ディリクレ分布モデルを採用する根拠として、ディリクレ分布の性質を2つ挙げる事ができる。まず1点目の性質として、ディリクレ分布は変数の和に対して良い性質を持っている。一般に、 $(k+1)$ 次元ディリクレ分布が与えられたとき、2個の変数の和を1個の変数とみなすことにより、 k 次元ディリクレ分布が得られる。例えば、上述したディリクレ分布の密度関数

$$\text{Dir}(x_1, \dots, x_{k-2}, x_{k-1}, x_k, y | \alpha_1, \dots, \alpha_{k-2}, \alpha_{k-1}, \alpha_k, \alpha_{k+1})$$

において、第 $(k-1)$ 変数と第 k 変数の和を考えると、新しい変数 $x' = x_{k-1} + x_k$ に対応するパラメータは2個の変数に対応するパラメータの和によって得られ、 k 次元ディリクレ分布の密度関数は

$$\text{Dir}(x_1, \dots, x_{k-2}, x', y | \alpha_1, \dots, \alpha_{k-2}, \alpha_{k-1} + \alpha_k, \alpha_{k+1})$$

と表される。よって、上述したディリクレ分布の変数の期待値の式を考慮すると、 k 次元ディリクレ分布の変数 x' の期待値は、元の $(k+1)$ 次元ディリクレ分布の変数 x_{k-1} と x_k の期待値の和に一致することが容易にわかる（期待値の加法性）。同様な性質が分散についても成立する。つまり、 k 次元ディリクレ分布の変数 x' の分散は、元の $(k+1)$ 次元ディリクレ分布の変数の和 $x_{k-1} + x_k$ の分散に一致する。

このディリクレ分布の変数の和に関する性質は、任意の2変数の組に対して成立する。さらにこの性質を繰り返し用いることにより、従属

変数 y を除くすべての変数の和

$$x_0 = x_1 + \dots + x_k$$

の期待値および分散は、独立変数が1個の場合のモデルであるベータ分布の期待値および分散と一致する(ただし、ベータ分布のパラメータをディリクレ分布のパラメータの和によって与えることとする)。本稿の多項分布・ディリクレ分布モデルにおいては、変数の和 x_0 の期待値および分散は事前分布と事後分布それぞれで、(事前分布)

$$E(x_0) = \frac{0}{0+},$$

$$V(x_0) = \frac{0}{(0+)^2(0+ + 1)},$$

(事後分布)

$$E(x_0) = \frac{0 + d_0}{0+ + p},$$

$$V(x_0) = \frac{(0 + d_0)(0+ + p - d_0)}{(0+ + p)^2(0+ + p + 1)},$$

(ただし、 $0+ = \sum_{i=1}^k$ とする)

と表されることがわかるが、これらの式は、全死因に対応する事前分布、事後分布の期待値および分散として解釈され、従来の市区町村別生命表で用いている二項分布・ベータ分布モデルによるベイズ推定において現れる式(事前分布については明示的な記述はないが、事後分布については平成12年市区町村別生命表²⁾に記述されている式)と一致している。ただし、報告書では、人口 p を作成基礎期間3年に対応するよう3倍していることを注意しておく。

以上の考察から、本稿で提案する多項分布・ディリクレ分布モデルによるベイズ推定は、事前分布のパラメータ $0+$ を適切に設定すれば、市区町村別生命表の結果と整合性のある結果になることがわかる。つまり、変数の和の期待値に関するディリクレ分布の性質から容易にわかるように、重複のないように分類した k 種類の死因に対する死因別中央死亡率(ベイズ推定値)の総和が全死因に対する中央死亡率に対応するので、全死因に対する中央死亡率(ベイズ推定値)が市区町村別生命表におけるベイズ

推定値と一致するように事前分布のパラメータを設定することが可能である。具体的な事前分布の与え方については、上記の注意を考慮しながら第 節(1-b)で述べる。また、期待値の加法性は、第 節(2)で死因分析の諸指標を算出する際に重要な役割を果たすことを注意しておく。

ディリクレ分布の性質として、次にパラメータの漸近挙動について挙げておく。一般に $(k+1)$ 次元ディリクレ分布

$$\text{Dir}(x_1, \dots, x_{k-1}, x_k, y \mid 1, \dots, k-1, k,)$$

が与えられたとき、パラメータが0に近づく分布は k 次元ディリクレ分布

$$\text{Dir}(x_1, \dots, x_{k-1}, y \mid 1, \dots, k-1,)$$

に法則収束(弱収束)する(漸近的性質)。定義域の異なる分布の収束であるので収束の解釈については注意を要するが、ここでは簡単に、対応するパラメータが0であるような変数は定数0として扱って良いと解釈する。

この点は第 節以降で具体的に事前分布を与えて死因別中央死亡率(ベイズ推定値)を算出しようとする際に問題になる。現実の観察においては、全国では少数の死亡が観測されるような死因について、小地域では全く観測されないことがある。特に市区町村別生命表では指標の地域差を保つために、ある市区町村の死亡率の事前分布の設定をする際に、全国の死亡状況を用いるわけではなく、地理的に近い地域の死亡状況によって事前分布を与える。そのため、死亡率の小さい死因について死亡が観測されず、死亡率の期待値

$$E(x_i) = \frac{i}{0+}$$

を0であると仮定する状況も生じることがある。このとき、事前分布は観測によれば0と仮定すべきパラメータ i が十分小さな値であるような $(k+1)$ 次元ディリクレ分布であると考え、計算上は k 次元ディリクレ分布として扱うこととする。このことは上述の漸近的性質から妥当である。

上述したディリクレ分布に関する期待値の加法性と漸近的性質より、本稿で扱う多項分布・

ディリクレ分布モデルで死因別中央死亡率（ベイズ推定値）を算出する際には、死因を重複しないように分類するとき、その分類数は自由に与えることができるといえる。つまり、本稿では市区町村別・性別・年齢階級別に死因別中央死亡率（ベイズ推定値）を算出するが、死因の分類数についてそれぞれの区分ごとに検討する必要はなく、すべての区分で同じ死因分類を用いることとする。

ベイズ推定値の作成方法

(1) ベイズ推定値の算出方法

本節では、市区町村別生命表における中央死亡率（ベイズ推定値）の算出方法を基にして、死因別中央死亡率（ベイズ推定値）の算出方法を説明する。

基礎データは平成16年から18年の3年間の人口動態統計および平成17年国勢調査とする。死因別中央死亡率（ベイズ推定値）を求めるために、人口動態統計の死亡数を死因別・市区町村別・性別・年齢階級別に集計し、年齢階級の区分については0歳、1～4歳、5～9、10～14、…、95～99、100歳以上とした。死因の分類については、従来の簡易生命表の15分類（再掲を含む）にならって、表1に示す過不足のない14分類を用いた。

1) 中央人口

まず、中央人口は作成基礎期間の中央にあたる平成17年7月1日現在の人口を国勢調査および人口動態統計調査を用いて推計する。本稿のベイズ推定においては、報告書²⁾に記載されている推計式を用いて中央人口を算出した（ただし、人口および死亡数の補正については行っていない）。

2) 事前分布

次に、死因別中央死亡率の事前分布を設定する。多項分布・ディリクレ分布モデルでは第節で述べたとおり、事前分布としてディリクレ分布を採用するが、そのパラメータの決定方法を述べる。まず、二次医療圏を基準にして全国をいくつかの地域に分割する。具体的には、平

表1 生命表上の死因分類

分類番号	死因分類名
1	悪性新生物
2	心疾患
3	脳血管疾患
4	肺炎
5	不慮の事故 ¹⁾
6	交通事故
7	自殺
8	肝疾患
9	腎不全
10	糖尿病
11	高血圧性疾患
12	結核
13	老衰
14	その他

注 1) 簡易生命表で用いる死因分類の場合は、不慮の事故に交通事故を含んでいるが、本稿では、死因を重複のないように分類するため、「不慮の事故（交通事故を除く）」と「交通事故」の2種類に分類した。

成12年市区町村別生命表の作成方法と同様に、人口規模の小さい二次医療圏については地理的に近いものを組み合わせて、年齢階級別死亡率が安定する程度の規模にまとめた。そして、分割した地域 A ごとに、性・年齢階級の死因別中央死亡率の事前分布を与えることとする。事前分布のパラメータ

$$i, i=1, \dots, 14 \text{ および}$$

を、以下の連立方程式により決定した。

$$\frac{i}{o_+} = E_i, i=1, \dots, 14$$

$$\frac{0}{(o_+)^2(o_+ + 1)} = V_0$$

ただし、

$$E_i = {}_c A \left\{ \frac{P^{(i)}}{{}_c A P^{(i)}} \frac{D_i^{(i)}}{3P^{(i)}} \right\}, i=1, \dots, 14$$

$$V_0 = {}_c A \left\{ \frac{P^{(0)}}{{}_c A P^{(0)}} \left(\frac{D_0^{(0)}}{3P^{(0)}} - \frac{{}_c A D_0^{(0)}}{3{}_c A P^{(0)}} \right)^2 \right\}$$

とする。ここで、 $P^{(i)}$ および $D_i^{(i)}$ はある市区町村 c における人口および第 i 死因による死因別死亡数であり、シグマ記号 ${}_c A$ は事前分布を設定する地域 A に属するすべての市区町村に対して和をとることを表す。よって、上記の数値 E_i および V_0 はそれぞれ、地域 A の死因別中央死亡率の期待値、地域 A の全死因中央死亡率

の分散になっていることがわかる。また、作成基礎期間の3年間に対応して中央人口を3倍して調整することに注意する。

上記の連立方程式を解けば、

$$E_i = E_i \left[\frac{E_0(1 - E_0)}{V_0} - 1 \right], i = 1, \dots, 14$$

$$= (1 - E_0) \left[\frac{E_0(1 - E_0)}{V_0} - 1 \right]$$

が得られる。ここで、 $E_0 = \sum_{i=1}^{14} E_i$ は、地域Aの全死因に対する中央死亡率であり、上記式を観察すればパラメータが市区町村別生命表における推定の際に設定した値と等しくなることがわかる。パラメータの和 $E_0 = \sum_{i=1}^{14} E_i$ についても同様であり、多項分布・ディリクレ分布モデルによるベイズ推定の際の事前分布が市区町村別生命表における二項分布・ベータ分布モデルによるベイズ推定の際の事前分布と整合性があることがわかる。

3) ベイズ推定値および分散

上記で求めたパラメータによって、市区町村ごとに性・年齢階級別の死因別中央死亡率

$$m_i, i = 1, \dots, 14,$$

の事前分布を与える。つまり、ディリクレ分布

$$\text{Dir}(m_1, \dots, m_{14} | y_1, \dots, y_{14},)$$

を事前分布とし、多項分布

$$\text{Pr}(D_1, \dots, D_{14} | m_1, \dots, m_{14})$$

$$= \frac{D_0!}{D_1! D_2! \dots D_{14}! (3P - D_0)!} m_1^{D_1} m_2^{D_2} \dots$$

$$m_{14}^{D_{14}} (1 - m_0)^{3P - D_0}$$

を尤度関数としてベイズ推定を行った結果得られる事後分布は、ベイズ推定の理論的説明で述べたとおり、

$$\text{Dir}(x_1, \dots, x_{14} | y_1 + D_1, \dots, y_{14} + D_{14}, + 3P - D_0)$$

となる。

上記の事後分布から、死因別中央死亡率（ベイズ推定値）および分散が以下に示す式のとおりに得られる。

$$E(m_i) = \frac{i + D_i}{o_i + 3P}$$

$$= \frac{o_i + 3P}{o_i + 3P} \frac{i}{o_i + 3P} + \frac{3P}{o_i + 3P} \frac{D_i}{3P}$$

$$V(m_i) = \frac{(i + D_i)(o_i + 3P - i - D_i)}{(o_i + 3P)^2(o_i + 3P + 1)}$$

$$= \frac{1}{o_i + 3P + 1} E(m_i)(1 - E(m_i))$$

上記の推定値の式からも、多項分布・ディリクレ分布による推定結果が二項分布・ベータ分布モデルの結果と整合性のあることがわかる。さらに、期待値 $E(m_i)$ が事前分布の期待値と観測値による粗死亡率 $D_i/3P$ との加重平均になっていることが観察される。

(2) 生命表の作成方法

ここで、(1)で解説した方法で求めた死因別中央死亡率（ベイズ推定値）の妥当性を検証するため、死因を除去した場合の平均余命および死因別死亡確率を試算する。生命表の作成方法は従来市区町村別生命表と同様に作成し、死因に関する扱いは従来簡易生命表と同様に行ったが、以下でそれぞれの方法の概略を述べる。

1) チャンの方法

年齢階級別の中央死亡率 ${}_n m_x$ が、(1)の方法で求まっているとする。中央死亡率から、死亡率 ${}_n q_x$ を

$${}_n q_x = \frac{n \cdot {}_n m_x}{1 + n(1 - {}_n a_x) {}_n m_x}, x < 95 \text{ かつ } q_{95} = 1$$

により計算する。ここで、 ${}_n a_x$ は死亡区間生存期間割合で、平成17年都道府県別生命表の数値から求めることができる²⁾。

上記の式により求めた死亡率から生存数 l_x を $l_0 = 100,000$ かつ $l_{x+n} = l_x(1 - {}_n q_x)$, $x < 95$

により計算し、再び死亡区間生存期間割合を用いて、定常人口 ${}_n L_x$ を

$${}_n L_x = n \cdot l_{x+n} + {}_n a_x (l_x - l_{x+n}) \cdot n, x < 95$$

かつ $L_{95} = \frac{l_{95}}{m_{95}}$

により計算する。

そして、定常人口 T_x を

$$T_x = {}_nL_x + {}_nL_{x+n} + \dots$$

により計算すると、平均余命 e_x° は

$$e_x^{\circ} = \frac{T_x}{l_x}$$

により求められる。

2) 特定死因の影響を除去した生命表

簡易生命表で用いている方法に従い、死因 i を除去した死亡率 ${}_nq_x^{(-i)}$ を算出し、1)で述べたチャンの方法で生命表を作成することにより、多項分布・ディリクレ分布モデルによる死因別中央死亡率(ベイズ推定値)の安定性について、第 節で考察する。死因を除去した死亡率 ${}_nq_x^{(-i)}$ の算出方法は、平成18年簡易生命表³⁾を要約すると、以下の式のようになる。

$${}_nq_x^{(-i)} \approx 1 - \exp \left\{ \left(1 - \frac{{}_nd_x^i}{{}_nd_x} \right) \log(1 - {}_nq_x) \right\}$$

ここで、 ${}_nd_x^i$ は通常の生命表における死因 i による死亡数である。簡易生命表では、観測された死亡数を用いて

$$\frac{{}_nd_x^i}{{}_nd_x} \approx \frac{D_i}{D_0}$$

により求めている(観測比による近似)が、本稿では多項分布・ディリクレ分布により求めた死因別中央死亡率(ベイズ推定値)を用いて

$$\frac{{}_nd_x^i}{{}_nd_x} \approx \frac{m_i}{m_0} \dots \dots \dots (式1)$$

により近似する(ベイズ比による近似)。上記の式から、死因 i を除去した死亡率を算出して生命表を作成した。死因 i を除去した死亡率 ${}_nq_x^{(-i)}$ は、通常の死亡状況と比較して死因 i の分だけ改善された状況における死亡率とみなすことができ、死因 i を除去した場合の平均余命 $e_x^{(-i)}$ と通常平均余命 e_x° との差(死因を除去した場合の平均余命の伸び)を観察することによって、死因 i が平均余命に与える影響を分析することが可能となる。

(補足) 多項分布・ディリクレ分布モデルによるベイズ推定値では加法性が成立するため、上記の近似式が適当である。一方、死因ごとに観察する場合は、特定の死因だけに注目して二項分布・ベータ分布モデルを適用することにより中央死亡率をベイズ推定することもできる⁴⁾。ただし、この方法では死因の加法性が成立せず、特定死因と全死因とを比較した場合に、不等式

$$m_i \leq m_0$$

が成立することを理論的に示すことができない。ディリクレ分布を導入すれば、上記のように死因を除去した死亡率を算出することができる。

3) 死因別死亡確率

上述の死因を除去した場合の平均余命の伸びとともに簡易生命表で行われている死因分析の指標である死因別死亡確率についても、本稿で求めた死因別中央死亡率をもとに算出する。

報告書³⁾によれば、 x 歳における死因別死亡確率 R_x^i の算定式は全死因に関する生命

図1-1 規模別の平均寿命の伸び(ベイズ比)

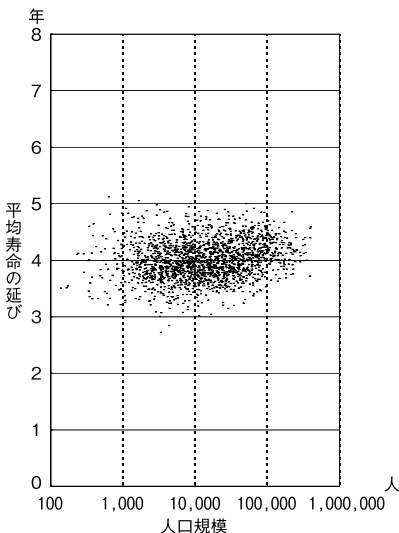


図1-2 規模別の平均寿命の伸び(確率比)

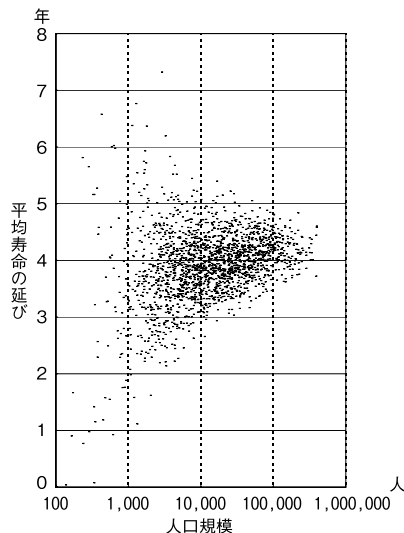


図2 人口10万以上の市の平均寿命の伸び

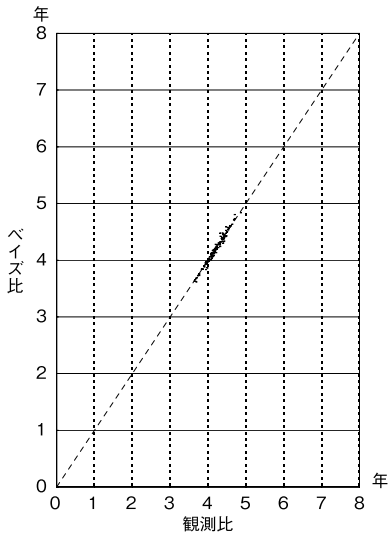


表2 平均寿命の伸び(ベイズ比)の分布

(単位 年)

	平均	標準偏差	参考 ²⁾
悪性新生物	4.0	0.3	4.10
心疾患	1.6	0.3	1.61
脳血管疾患	1.2	0.2	1.15
肺炎	1.0	0.2	0.95
不慮の事故 ¹⁾	0.8	0.2	0.66
交通事故(再掲)	0.3	0.1	0.27
自殺	0.9	0.2	0.79
肝疾患	0.2	0.1	0.27
腎不全	0.2	0.0	0.17
糖尿病	0.1	0.0	0.15
高血圧性疾患	0.0	0.0	0.06
結核	0.0	0.0	0.05
その他	2.2	0.3	-
3大死因(再掲)	8.8	0.8	8.65

注 1) 交通事故を含む
2) 平成17年都道府県別生命表における全国の伸び

図3-1 地域別の平均寿命の伸び(ベイズ比)

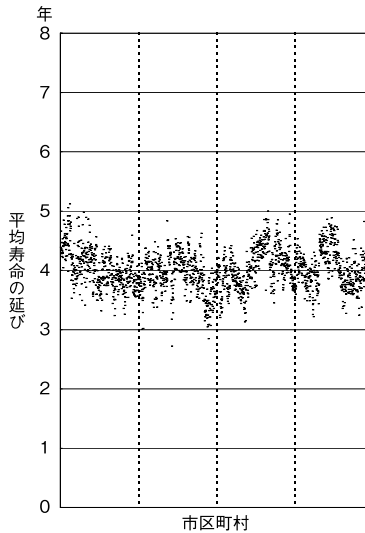
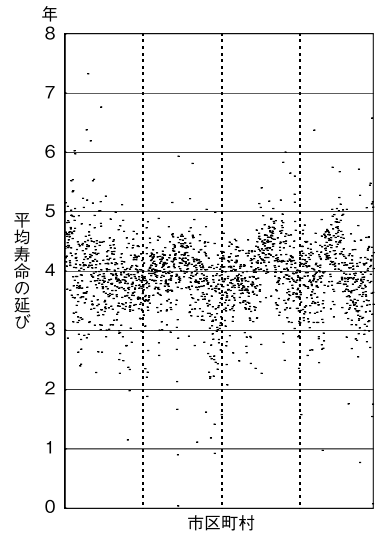


図3-2 地域別の平均寿命の伸び(確率比)



表における死亡数

${}_n d_x^i$ を用いて,

$$R_x^i = \frac{\int_{\geq x} {}_n d_x^i}{l_x}$$

$$= \frac{1}{l_x} \int_{\geq x} \left(\frac{{}_n d_x^i}{{}_n d_x} {}_n d_x \right)$$

と表せる。死因を除去した場合の平均余命の伸びを算出したときと同様に死亡数の比が現れるので、本稿では式1により得られるベイズ比によ

り、死亡数の比を近似して死因別死亡確率を試算した。

死因別死亡確率 R_x^i は、生命表の上で年齢 x 歳の者が死因 i によって死亡する確率を表した指標である。生命表の上ではすべての者が何らかの死因で死亡することになるので、すべての死因について死因別死亡確率の和を求めれば常に1に一致することが期待される。本稿の方法による死因別中央死亡率(ベイズ推定値)は死因に関する加法性が成立するので、死因別死亡

確率 R_x^i の和が1に一致する。

結 果

本稿では多項分布・ディリクレ分布モデルを導入することによって、死因別中央死亡率(ベイズ推定値)を算出した。この結果を検証するため、死因別中央死亡率の加法性を利用して、死因を除去した場合の平均余命の伸びと死因別死亡確率を試算した結果を示す。

(1) 死因を除去した場合の平均余命の伸び

はじめに、本稿の方法によって得られる小地域に対する指標の安定性を確認するために、市区町村の人口規模と死因を除去した場合の平均余命の伸びとの関係を図1-1, 1-2に示す。除去した死因は死因1(悪性新生物)であり、男0歳の平均余命、つまり平均寿命の伸びである。比較として、第節(2)2)で述べた観測比による近似とベイズ比による近似の2通りの方法で平均余命の伸びを算出した。

図1-1, 1-2から、人口規模の小さな市区町村に対して、ベイズ比による近似によって算出した結果は観測比による近似によって算出した結果よりばらつきが小さい。このことは、観測比による方法では偶然変動の影響を受けてしまうが、本稿の方法によってその影響を排除することができたと考えられる。一方、人口規模の大きな市区町村に対しては、ベイズ比による近似と観測比による近似とでほぼ等しい結果となっていることが図から読み取れる。さらに

図4-1 規模別の死因別死亡確率(ベイズ比)

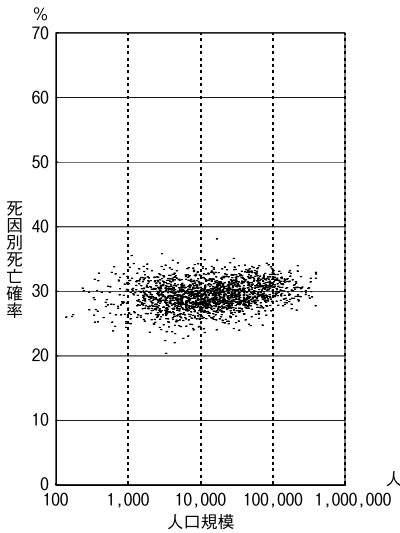


図4-2 規模別の死因別死亡確率(確率比)

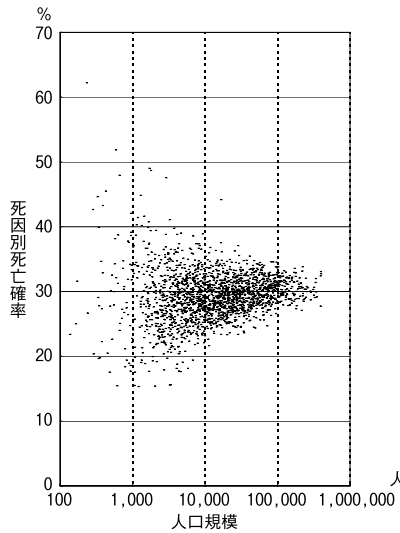


図5-1 地域別の死因別死亡確率(ベイズ比)

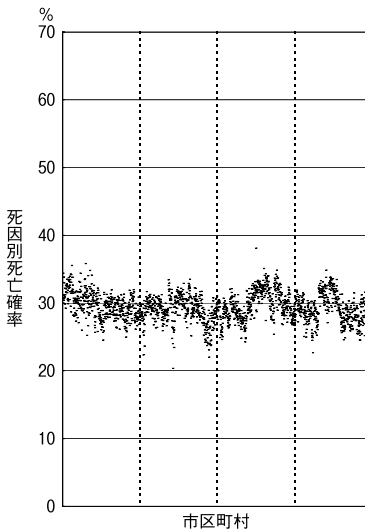
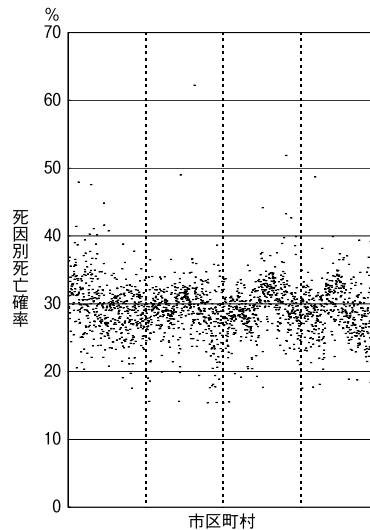


図5-2 地域別の死因別死亡確率(確率比)



詳しくみるために男性の人口が10万以上の市についてベイズ比と観測比を比較した場合の平均寿命の伸びの散布図を図2に示す。

また、多項分布・ディリクレ分布モデルによる死因を除去した場合の平均余命の伸びが地域間の差を表す指標となることを確認するため、事前分布を設定した地域の順に並べた算出結果を図3-1, 3-2に示す。横軸に北から南へ二次医療圏ごとに市区町村を並べて図示している。図から、おおむね平均的な値の近くに分布しているが、地域的な傾向を持って分布していることが観察される。さらに、その傾向の中でも各市

表3 死因別死亡確率（ベイズ比）の分布

（単位 %）

	平均	標準偏差	参考 ²⁾
悪性新生物	29.4	2.1	30.13
心疾患	14.8	1.7	14.78
脳血管疾患	11.7	1.8	11.54
肺炎	12.3	1.6	12.35
不慮の事故 ¹⁾	4.1	0.9	3.75
交通事故（再掲）	1.1	0.4	0.91
自殺	2.9	0.7	2.62
肝疾患	1.3	0.4	1.41
腎不全	1.9	0.5	1.92
糖尿病	1.1	0.3	1.08
高血圧性疾患	0.4	0.3	0.45
結核	0.2	0.2	0.29
老衰	2.3	1.1	2.10
その他	17.6	1.8	-

注 1) 交通事故を含む

2) 平成17年都道府県別生命表における全国の伸び

区町村の数値はばらつきをもって分布していることから、地域の特性を表した指標が得られたと考えることもできる。

死因1以外の死因を除去した場合についても作成した指標の、全国の市区町村における平均余命の伸びの分布から平均値と標準偏差を算出した。表2に死因を除去した場合の男の平均寿命の伸び（ベイズ比）の場合の結果を示し、参考として平成17年都道府県別生命表における全国の平均寿命の伸びも記載した。分布の平均値は人口規模を考慮していないため参考数値と一致するわけではないが、両者の値はおおむね近くなっている。ここで死因の分類のうち、不慮の事故については都道府県別生命表の表章に合わせて交通事故を含むものとしていることに注意する。

（2）死因別死亡確率

第節（2）3）で説明した死因別死亡確率の算出結果からも（1）と同様な考察が得られた。死因1による死因別死亡確率（男，0歳）についての結果を図4-1，4-2，5-1，5-2および表3に示す。ただし，死因別死亡確率（観測比）の数値については，死亡確率の基準である全死因中央死亡率はベイズ推定した値により算出しているために，死亡が観測されない年齢階級について死亡確率がそれぞれの死因に配分されることに注意する。

また，死因別死亡確率については，本稿の算

表4 ある市区町村における死因別死亡確率

（単位 %）

	観測比	ベイズ比 ²⁾
悪性新生物	15.6	20.4
心疾患	16.1	16.1
脳血管疾患	24.9	21.1
肺炎	12.5	13.2
不慮の事故 ¹⁾	1.6	2.0
交通事故（再掲）	0.3	0.5
自殺	3.6	2.8
肝疾患	0.5	0.9
腎不全	1.1	1.4
糖尿病	2.3	1.9
高血圧性疾患	0.4	0.4
結核	0.4	0.3
老衰	1.7	1.6
その他	19.3	17.6

注 1) 交通事故を含む

2) 端数処理の関係で再掲を除く合計が100に一致しない。

出方法によれば死因に関する加法性が成立するので，重複のないようなすべての死因に対する死因別死亡確率の和が1に一致する。端数の処理の関係で合計が一致しないこともあるが，ある市区町村の死因別死亡確率について和がほぼ1になることを表4で確認している。ここでも，不慮の事故については交通事故を含む。

考察と注意

本稿では全死因に対する中央死亡率（ベイズ推定値）と整合性を持つ死因別中央死亡率（ベイズ推定値）の算出方法を提案した。この方法は，全死因については市区町村別生命表で用いられている方法と整合性を持つ方法であって，事前分布の設定の際に特定の市区町村の周辺の死亡状況を用いる方法であることから，経験的ベイズ推定法となっている。また本稿では，その方法を検証するために死因を除去した場合の平均余命の伸びと死因別死亡確率を試算した。

本稿の方法によって得られる指標は，第節でも述べたように，出現数の少なさに起因する偶然性の影響を減少させた数値となっていると考えられ，さらに死因に対する加法性が保たれており従来の全死因に対する指標と矛盾しない結果が得られる。以上のことから，多項分布・ディリクレ分布モデルが地域間の比較を行う際

の有用な資料を作成するという観点から今後継続して検討する価値のある方法であると考えます。また、本稿の指標の特徴として、死因の分類の扱いやすさも指摘しておく。特定の死因についての指標を得るためには、基礎数値として特定の死因およびその他の死因の2通りの分類で本稿の方法を適用すれば良い事が、ディリクレ分布の性質から導かれる。つまり、その他の死因については分類の仕方を検討する必要がなく基礎数値の扱いが容易になるという特徴がある。

一方で今回の結果は、作成基礎期間が1期間でありベイズ推定の事前分布の与え方についても1通りの方法であるので、今後さまざまな設定に対して指標を試算して検討を重ねることが望まれる。さらに、全死因に対する指標としては生命表以外にも標準化死亡比や年齢調整死亡率などの指標が知られているので、これらの指標についても本稿と同様なモデルを適用して死因分析を行うことが期待される。特に年齢調整

死亡率については性・年齢階級別の死亡率を多項分布・ディリクレ分布モデルを適用することが考えられる。

最後に、本文中の所見は筆者の個人的な見解であることを申しあげておく。

文 献

- 1) 平子哲夫, 佐伯則英, 中田正. 人口動態市区町村別統計へのベイズ統計の応用について(1)標準化死亡比への応用. 厚生指標 1999; 46(10): 3-11.
- 2) 厚生労働省大臣官房統計情報部. 平成12年市区町村別生命表 2003: 12-4.
- 3) 厚生労働省大臣官房統計情報部. 平成18年簡易生命表 2007: 16.
- 4) 鈴木雪夫. 小地域における死亡指標 - ベイズ統計学からのアプローチ. 平成元年度厚生科学研究報告書 1990.

付録 補足

以下では、第 3 節において述べたパラメータの漸近挙動について正確に記述することにする。定義 $n - 1$ 次元空間 R^{n-1} 上に以下の式で定義される関数 $f_n(x_1, \dots, x_n)$ を密度関数とする分布を n 次元ディリクレ分布と呼び、記号 $D_n(x_1, \dots, x_n)$ で表す。

$$f_n(x_1, \dots, x_n) = \begin{cases} \frac{1}{B(\alpha_1, \dots, \alpha_n)} x_1^{\alpha_1-1} x_2^{\alpha_2-1} \dots x_n^{\alpha_n-1} & \text{if } x_1 + x_2 + \dots + x_{n-1} < 1, x_i > 0 \quad (i) \\ 0 & \text{otherwise} \end{cases}$$

ただし、 $x_n = 1 - (x_1 + x_2 + \dots + x_{n-1})$ とし、 R^{n-1} 上には標準的な測度が定義されているとする。

ディリクレ分布 $D_n(x_1, \dots, x_n)$ の定義に現れる x_n を別の変数とみなすことで n 次元空間上の分布が自然に得られる。つまり、空間 R^n 上の $n - 1$ 次元空間から n 次元空間への写像 i_{n-1} を

$$(x_1, x_2, \dots, x_{n-1}) \mapsto (x_1, x_2, \dots, x_{n-1}, 1 - (x_1 + x_2 + \dots + x_{n-1}))$$

で定義したときに、分布 $D_n(x_1, \dots, x_n)$ の写像 i_{n-1} による像分布

$$(i_{n-1})_* D_n(x_1, \dots, x_n)$$

が空間 R^n 上に定義される。

さらに1次元低い $n - 2$ 次元空間 R^{n-2} 上に定義される分布 $D_{n-1}(x_1, \dots, x_{n-1})$ を考えると、合成写像 $i_{n-1} \circ i_{n-2}$ による像分布として、空間 R^n 上に定義された別の分布

$$(i_{n-1} \circ i_{n-2})_* D_{n-1}(x_1, \dots, x_{n-1})$$

が得られる。

主張 パラメータ α_n が 0 に近づくとき、 n 次元ディリクレ分布の像分布は $n - 1$ 次元ディリクレ分布の像分布に弱収束する。

証明 n 次元ディリクレ分布 $D_n(x_1, \dots, x_n)$ の写像 i_{n-1} による像分布を μ_n とかき, $n-1$ 次元ディリクレ分布 $D_{n-1}(x_1, \dots, x_{n-1})$ の写像 $i_{n-1}i_{n-2}$ による像分布を μ_{n-2} とかくことにする。主張を示すために, 以下のことを示す。

$$(*) \quad \text{任意の開集合 } G \subset \mathbb{R}^n \text{ に対して, } \liminf_k \mu_k(G) \geq \mu(G)$$

ここで, 上記の不等式を変形して,

$$\begin{aligned} \liminf_n \mu_n(G) &\geq \mu(G) \\ &> 0, \quad A > 0, 0 < \mu_n < A, \\ \inf_{0 < \mu_n < A} \mu_n(G) &\geq \mu(G) - \\ &> 0, \quad A > 0, 0 < \mu_n < A, \\ \mu(G) &\geq \mu(G) - \end{aligned}$$

また, n 次元空間から $n-1$ 次元空間への射影 p を

$$(x_1, x_2, \dots, x_n) \mapsto (x_1, x_2, \dots, x_{n-1})$$

により定義し, $n-1$ 次元ディリクレ分布 $D_{n-1}(x_1, \dots, x_{n-1})$ を定義する密度関数の台を μ_{n-2} で書くことにする。つまり

$$\mu_{n-2} = \{(x_1, x_2, \dots, x_{n-1}) \mid x_1 + x_2 + \dots + x_{n-1} < 1, x_i > 0 (i)\}$$

である。

いま, 任意の開集合をひとつ固定する。

$$I = G \cap i_{n-1}i_{n-2}(\mu_{n-2})$$

$$\tilde{I} = p^{-1}p(I)$$

また, 正の数 $\epsilon > 0$ に対して,

$$B = \{(x_1, x_2, \dots, x_n) \mid \mu_n < x_n < \mu_n + \epsilon\}$$

$$J = B \cap G \cap \tilde{I}$$

$$\tilde{J} = p^{-1}p(J)$$

すると,

$$\begin{aligned} \mu(G) &\geq \mu(G \cap \tilde{I}) = \mu(\tilde{I}) - \mu(G^c \cap \tilde{I}) \\ &= \mu(\tilde{I}) - \mu(G^c \cap B \cap \tilde{I}) - \mu(G^c \cap B^c \cap \tilde{I}) \\ &\geq \mu(\tilde{I}) - \mu(\tilde{J}) - \mu(B^c \cap \tilde{I}) \end{aligned}$$

よって, 任意の正の数 ϵ に対して, 正の数 δ と η を十分小さくとれば,

$$\mu(G) \geq \mu(\tilde{I}) - \delta = \mu(G) - \delta$$

が成り立つから, 弱収束することが示せた。(証明終)