

一般口演 | 医療データ解析

一般口演7

医療データ解析

2019年11月22日(金) 14:50 ~ 16:20 D会場 (国際会議場 2階中会議室201)

[2-D-2-02] ベイズ統計学を用いた死亡率の推定手法に関する考察—大規模医療データベースの分析結果に基づく事前分布の設定

○古畑 宏樹^{1,2}、小川 泰右¹、松尾 亮輔¹、山崎 友義¹、荒木 賢二¹ (1. 宮崎大学医学部附属病院 病院IR部, 2. 宮崎大学大学院 医学獣医学総合研究科)

キーワード : Bayesian statistics, Conjugate prior distribution, Decision tree, Electronic medical record

【目的】ベイズ統計学を用いた死亡率の推定事例として、小地域推定という手法を用いて市町村単位の値を求めたものが多い。この手法は、地域単位でのサンプルサイズが小さい点などを克服するための様々な工夫がなされているが、研究成果の積み上げの結果としてモデリングの方法が複雑化している。本研究では、先に提案した「直近入院時の死亡確率を過去の治療状況を用いて予測するモデル」を事前分布として使用し、モデリングの単純化可能性について考察した。

【対象と方法】宮崎大学医学部附属病院の電子カルテに蓄積されているデータを使用した。事前確率の予測モデル（決定木）の構築には入院年月日が2013年4月1日から2015年3月31日までのもの（N=22,424、うち死亡831）を、検証用データには入院日が2015年4月1日から2018年3月31日までのもの（N=22,658、うち死亡802）を使用した。

【結果】検証用データにおける粗死亡率（更新前）と更新後死亡率（事後確率、共役事前分布に基づき計算）は、それぞれ $3.54\% \pm 0.12\%$ 、 $3.62\% \pm 0.09\%$ であった（平均値 \pm 標準偏差）。すべてのノードについて、更新後死亡率の標準偏差が粗死亡率を下回っていた。

【結語】更新後死亡率の標準偏差が粗死亡率のそれを大きく下回るのは、先行研究にも同様の結果が多くあり、事後確率への更新により値のばらつきを抑制できる点において利点があることがわかる。将来における本研究成果の応用として、冒頭に述べた「小地域ごとの推定」を「医療圏ごとの推定」に読み替え、医療圏ごとの基幹病院（ \equiv DPC対象病院）主体による情報交換ネットワークを構築しデータの蓄積とモデルの個別化を行うことで、医療政策面における貢献可能性が示唆される。

ベイズ統計学を用いた死亡率の推定手法に関する考察 —大規模医療データベースの分析結果に基づく事前分布の設定—

古畑 宏樹^{*1,*2}、小川 泰右^{*1}、松尾 亮輔^{*1}、山崎 友義^{*1}、荒木 賢二^{*1}

*1 宮崎大学医学部附属病院 病院 IR 部、*2 宮崎大学大学院 医学獣医学総合研究科

Estimation of Mortality Rate by Bayesian Statistics - Construction of Prior Distribution from Data Analysis of Large Size Electronic Medical Record-

Hiroki Furuhata^{*1,*2}, Taisuke Ogawa^{*1}, Ryosuke Matsuo^{*1}, Tomoyoshi Yamasaki^{*1}, Kenji Araki^{*1},

*1 Department of Hospital Institutional Research, University of Miyazaki Hospital, *2 Graduate School of Medicine and Veterinary Medicine, University of Miyazaki

This study has suggested a method of estimation of mortality rate within risk adjustment based on the conjugate prior distribution, one of the most important methods in the Bayesian statistics. The prior probability was estimated from the decision tree constructed by information collected at previous hospitalization. The posterior probability was calculated by the conjugate prior distribution of beta distribution. The cohort for validation includes 22,658 patients (802 patients has died) from the electronic medical records of the University of Miyazaki Hospital. The period of data collection was three years (from 1 April 2015 to 31 March 2018). As a result, the difference of mean between crude mortality rate and posterior probability of mortality rate was very little. In addition, standard deviation of the posterior probability was lower than the crude mortality rate. Therefore, this method can decrease an error of target indicator, and would improve an evaluation method of service quality by wide introduction of the method of estimation of the posterior probability.

Keywords: Bayesian statistics, Conjugate prior distribution, Decision tree, Electronic medical record.

1. 目的

ベイズ統計学を用いた死亡率の推定事例として、小地域推定 (Small Area Estimation) と呼ばれる方法により市町村単位の値を求めたものが多い。この手法は、地域単位でのサンプルサイズが小さい点などを克服するための様々な工夫がなされているが、研究成果の積み上げの結果としてモデリングの方法が個別化・細分化されている。本研究では、保健医療統計の品質向上などへの貢献を念頭に、先に提案した「直近入院時の死亡確率を過去の治療状況を用いて予測するモデル」²⁾を、ベイズ統計学における事前分布として使用することで、モデリングの汎用化の可能性について検討した。

2. 方法

2.1 データ準備

分析用データベースは、宮崎大学医学部附属病院の電子カルテに蓄積されているデータを使用した。事前確率の予測モデルの構築には入院年月日が2013年4月1日から2015年3月31日までのもの(N=22,424、うち死亡831)を、検証用データには入院日が2015年4月1日から2018年3月31日までのもの(N=22,658、うち死亡802)を使用した。

2.2 予測モデルを用いた事前確率の設定

事前確率の予測モデルは性別、直近入院時の年齢、過去の入院回数、過去入院時における手術の有無を説明変数とする決定木とした。なお、決定木の適用結果として与えられる予測死亡率は、ノード別の粗死亡率 (Crude mortality rate: CMR) に相当する。

No.	過去の入院	性別	年齢	過去の入院回数	過去の手術	合計患者数	死亡患者数	PMR (%)
1	なし	男性	≥85	NA	NA	416	48	11.54
2	なし	女性	≥85	NA	NA	507	54	10.65
3	なし	男性	0 or 75-84	NA	NA	1642	102	6.21
4	なし	男性	65-74	NA	NA	1865	89	4.77
5	なし	女性	75-84	NA	NA	1358	62	4.57
6	なし	女性	0 or 65-74	NA	NA	1667	50	3.00
7	なし	男性	20-64	NA	NA	3306	119	3.60
8	なし	男性	1-19	NA	NA	1160	12	1.03
9	なし	女性	20-64	NA	NA	3640	46	1.26
10	なし	女性	1-19	NA	NA	693	8	1.15
11	あり	NA	≥3	≥3	なし	569	56	9.84
12	あり	NA	NA	≥3	あり	493	33	6.69
13	あり	NA	≥65	2	なし	315	22	6.98
14	あり	男性	65-74 or ≥85	1	なし	474	26	5.49
15	あり	男性	75-84	1	なし	312	11	3.53
16	あり	女性	NA	1	なし	585	19	3.25
17	あり	NA	≥75	1 or 2	あり	347	9	2.59
18	あり	NA	65-74	1 or 2	あり	339	6	1.77
19	あり	男性	<65	1 or 2	NA	984	32	3.25
20	あり	男性	<65	1 or 2	NA	329	4	1.22
21	あり	女性	<65	2	NA	345	7	2.03
22	あり	女性	<65	1	NA	1078	16	1.48

図1 決定木の構造
(参考文献²⁾の Fig.1 を改変)

2.3 事後確率の計算

本研究では、事前分布の従う確率分布は二項分布と定め、事後確率を共役事前分布 (ベータ分布) に基づく平均・分散の算式を使用し計算した。

a、b、c、n をそれぞれ事前分布における死亡者数、事前分布における生存者数、検証用データにおける死亡者数、検証用データにおける全患者数とおくと、事後確率 (Bayesian adjusted mortality rate: BAMR) の平均値は(a+c)/(a+b+n)、分散は(a+c)(b+n-c)/(a+b+n)²(a+b+n+1)となる。

3. 結果

3.1 患者背景

検証用データの患者背景として、図1に示した予測モデルの説明変数の分布を示す。

表1 患者背景

		合計 患者数	死亡 患者数	CMR (%)
全患者		22658	802	3.54
過去の入院	あり	16456	562	3.42
	なし	6202	240	3.87
性別	男性	11890	522	4.39
	女性	10768	280	2.60
年齢	平均値(標準偏差)	55.1 (24.3)		
	中央値(最小値, 最大値)	63 (0, 103)		
	0	520	19	3.65
	1-19	2317	32	1.38
	20-64	9357	245	2.62
	65-74	5196	194	3.73
	75-84	4061	221	5.44
	over 84	1207	91	7.54
過去の入院回数	平均値(標準偏差)	0.5 (1.2)		
	中央値(最小値, 最大値)	0 (0, 25)		
	0	16456	562	3.42
	1	3793	86	2.27
	2	1222	59	4.83
	3以上	1187	95	8.00
過去の手術	あり	3509	146	4.16
	なし	2693	94	3.49

3.2 CMRとBAMRの比較

検証用データにおける粗死亡率(ペイズ更新前)と更新後死亡率(事後確率)は、それぞれ 3.5%±18.5%、3.6%±0.1%であった(平均値±標準偏差)。決定木のノード別に比較した場合も、粗死亡率と更新後死亡率との間で平均値の差が小さく、標準偏差の差が大きい(更新後死亡率のほうが小さい)傾向は変わらなかった。

表2 CMRとBAMRの比較

ノード 番号	患者数		平均 (%)		標準偏差 (%)	
	合計	死亡	CMR	BAMR	CMR	BAMR
1	454	40	8.81	10.11	1.33	1.02
2	501	38	7.58	9.13	1.18	0.91
3	1774	117	6.60	6.41	0.59	0.42
4	1974	83	4.20	4.48	0.45	0.33
5	1361	67	4.92	4.74	0.59	0.41
6	1781	44	2.47	2.73	0.37	0.28
7	3266	107	3.28	3.44	0.31	0.22
8	1097	11	1.00	1.02	0.30	0.21
9	3542	46	1.30	1.28	0.19	0.13
10	706	9	1.27	1.22	0.42	0.29
11	285	24	8.42	9.37	1.64	1.00
12	902	71	7.87	7.46	0.90	0.70
13	182	15	8.24	7.44	2.03	1.18
14	314	13	4.14	4.95	1.12	0.77

15	192	10	5.21	4.17	1.60	0.89
16	412	9	2.18	2.81	0.72	0.52
17	686	21	3.06	2.90	0.66	0.52
18	739	23	3.11	2.69	0.64	0.49
19	549	16	2.91	3.13	0.72	0.44
20	615	21	3.41	2.65	0.73	0.52
21	308	8	2.60	2.30	0.90	0.59
22	1018	9	0.88	1.19	0.29	0.24
ALL	22658	802	3.54	3.62	0.12	0.09

3.3 数学的検証

前節に示した標準偏差の変化について、ベータ分布を共役事前分布した場合における事前分布と事後分布の分散の算式の差を取り、変化の妥当性について検証した。なお、以下に示す式における m, n, x, y は、それぞれ患者数(決定木の構築データ)、データ(検証用データ)、死亡患者数(決定木の構築データ)、死亡患者数(検証用データ)を指し、m と n は x と y よりも非常に大きな値を取るものとする。

<分散の差(分子)>

$$m^3(2nx+2xy-my-ny)+m^2y(2x+y)+m^3y(y-1)+m^2nx(n-3x)+m^2n(nx-y)+n^2x(m-x)(m+n+1)+m^2x(m-2x) (\Rightarrow > 0)$$

<分散の差(分母)>

$$m^2(m+x)(m+n)^2(m+n+1) (\Rightarrow > 0)$$

以上の結果から、少なくとも分散については事後分布の値が事前分布の値より小さくなり、BAMR への「更新」に伴うばらつきが抑制が図られることが示唆される。

4. 考察

4.1 推定手法の評価

更新後死亡率の標準偏差が粗死亡率のそれを大きく下回るのは、先行研究にも同様の結果が多くあり、事後確率への更新により値のばらつきを抑制できる点において利点があることがわかる。これは、3.3 節における式の変形結果から、使用データ数が非常に大きい条件下において証明されており、頑健性の高い手法であると言える。

共役事前分布を利用して死亡率の推定精度を向上させている事例として、厚生労働省が作成している市区町村生命表が³⁾あげられる。これに比べ、本研究で提案する手法は、事前確率を患者背景ごとに細かく設定している点において優れている。また、事前確率の推定モデルに決定木を採用していることから、患者の層別化を効率的に行うことができる。

4.2 本研究の応用可能性

医療機関の機能分化に伴い複数の医療機関の受診が進んだ結果、医療情報の利活用においては患者の状態を切れ目なく追跡することが重要となっている。例えば地域医療の枠組みの中で医療情報の連携(仮称:医療情報共有ネットワーク)を図り、どの病院においても事前確率を推定できるようにすることで、医療の質をより精緻に評価可能にすることが期待される。

図2に前段落にて述べた医療情報連携および事前確率の活用の概念図を示す(I:情報、M:金銭、S:サービス)。まず、医療サービスの受け手である患者は、医療機関に M₁、加入健保機関に M₂を支払うことにより、医療サービス S の提供を受ける。一方、医療機関は請求情報として I₁を健保機関に送信することにより、健保機関から M₃を受け取る。なお、これらの間に下式に示す関係を仮定すると、S-F の絶対値が小さいほど、M₂の徴収(による所得の再分配)を通じた医療サービ

スの平等な提供が効率的に行われていると評価できる。

$S=M_2+M_3$ (医療機関の収入)

$F=M_1+M_2$ (患者の医療負担金)

$S-F=M_2-M_3$ (健保機関の手持ちキャッシュ)

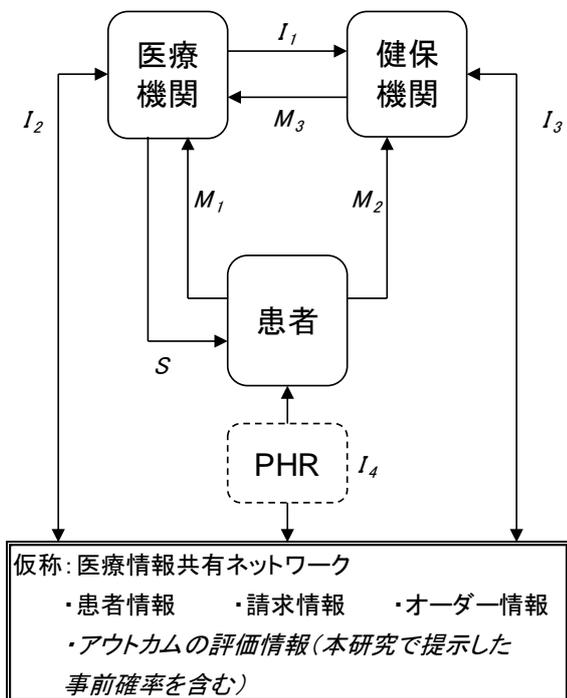


図2 サービス・金銭・情報の流通フロー

これらの金銭のやり取りの証跡として、医療機関と健保機関は I_2 と I_3 による情報の登録・参照を行うものとする。また、患者も PHR(Personal Health Record)を経由し追加情報の登録や治療履歴の参照を行い、情報量の充実を図るものとする。

参考文献

- 1) Ghosh M, Rao JNK. Small Area Estimation: An Appraisal. Stat Sci, 1994;9:55-93
- 2) 古畑宏樹, 小川泰右, 荒木賢二. 「リスク調整後臨床指標」算出のための試行分析—データ・クラスタリングによる患者リスクの比較. 第22回医療情報学春季学術大会, 2018年.
- 3) 厚生労働省「<参考>ベイズ推定とは」
[<https://www.mhlw.go.jp/toukei/saikin/hw/life/ckts15/dl/ckts15-05.pdf>] (2019年8月13日閲覧)