

区分指数モデル及びフレイルティモデルに対する

ベイズ流生存時間解析

○矢田 真城¹ 魚住 龍史^{1,2} 浜田 知久馬¹

¹東京理科大学大学院 工学研究科 経営工学専攻

²京都大学大学院医学研究科 医学統計生物情報学

Bayesian survival analysis for piecewise exponential and frailty models

Shinjo Yada¹, Ryuji Uozumi^{1,2}, and Chikuma Hamada¹

¹*Department of Management Science, Graduate School of Engineering, Tokyo University of Science*

²*Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine*

要旨

SAS では、比例ハザードモデルによる生存時間解析を行うためのプロシジャとして、PHREG プロシジャが用意されている。Ver.9.2 より BAYES ステートメントが追加され、比例ハザードモデルによるベイズ流生存時間解析が可能となった。また、Ver.9.3 より RANDOM ステートメントが追加され、フレイルティモデルによる生存時間解析が可能となった。フレイルティモデルは、非独立な複数の生存時間データを解析するために、比例ハザードモデルを拡張したものである。比例ハザードモデルに対するベイズ流解析では、ハザードには事前分布を想定せず、部分尤度に基づいてパラメータ推定値を算出するアプローチと、ハザードにも事前情報を付与してパラメータ推定値を算出するアプローチがある。本稿では、区分指数モデルとフレイルティモデルに対するベイズ流生存時間解析についてまとめ、SAS による適用例を報告する。

キーワード：ベイズ流解析，生存時間解析，比例ハザードモデル，区分指数モデル，フレイルティモデル，PHREG プロシジャ，MCMC プロシジャ

1. はじめに

ある基準の時刻からある目的の反応（観測対象とする個体に対し一度だけ非再起的に起きる事象であるとする）が起きるまでの時間を解析対象とする解析手法の総称を、生存時間解析とよぶ（大橋・浜田, 1995; 大橋ら, 2016）。

生存時間解析においてよく用いられる手法のひとつとして、Cox (1972)の比例ハザードモデルによる解析があげられる。比例ハザードモデルは、生存時間（基準ハザード関数）に特定の分布を仮定することなく共変量の効果を推定することから、セミパラメトリックなモデルとよばれる。SAS では、比例ハザードモデルを

行うためのプロシジャとして、PHREG プロシジャが用意されている。Ver.9.2 より、PHREG プロシジャに BAYES ステートメントが追加され(SAS Institute Inc., 2008)、比例ハザードモデルを用いて、ベイズ流に生存時間解析を行うことが可能となった。

比例ハザードモデルを用いてベイズ流に解析する場合、ハザード関数に特定の分布を仮定せず、回帰係数に事前分布を想定し、部分尤度に基づき共変量の効果を推定するアプローチがある(矢田ら, 2016)。これとは別に、ハザード関数を複数のパラメータでモデル化し、回帰係数及びハザード関数に対して事前情報を付与した上で、推論を行うアプローチがある。後者に属するアプローチのひとつとして、時間軸を幾つかの区間に分割し、各区間でハザードが一定であるとみなした、区分指数モデル(piecewise exponential model)による解析手法がある。

また、生存時間解析において、比例ハザードモデルのほかに用いられるセミパラメトリックなモデルとして、フレイルティモデル(frailty model)がある。フレイルティモデルは、非独立な複数の生存時間データを解析するために、比例ハザードモデルを拡張させたモデルである。SAS では、Ver.9.3 より、PHREG プロシジャに RANDOM ステートメントが追加され(SAS Institute Inc., 2011)、フレイルティモデルを用いた解析が実行可能となった。フレイルティモデルに対するベイズ流生存時間解析では、クラスターのフレイルティの分布を規定するパラメータに事前分布を想定し、部分尤度を用いて推論を行う。別のアプローチとして、ハザードをモデル化するパラメータにも事前分布を想定し、これらの同時分布を用いて推論を行う方法がある。

本稿では、区分指数比例ハザードモデルとフレイルティモデルに対するベイズ流生存時間解析をまとめ、SAS での適用例を紹介する。区分指数比例ハザードモデルについては、ハザード関数をモデル化するパラメータと回帰係数に事前分布を想定し、モデルパラメータ推定値を算出する方法を説明する。フレイルティモデルについては、回帰係数とフレイルティに事前分布を想定し、部分尤度に基づきモデルパラメータ推定値を算出する方法を説明する。最後に、フレイルティモデルについて、回帰係数とフレイルティに加えてハザードにも事前情報を与え、これらの同時確率分布を用いてモデルパラメータを推定する方法を示す。

2. 区分指数モデル

以下、解析対象となる時間を生存時間とよび、対象とする事象をイベントとよぶことにする。

i 番目の被験者($i=1,2,\dots,n$)に対して、イベント発現時間あるいは打ち切り時間を t_i 、打ち切りを表す変数を v_i (イベント発現のとき $v_i=1$, 打ち切りのとき $v_i=0$)、 p 個の共変量を $\mathbf{x}_i=(x_{i1}, x_{i2}, \dots, x_{ip})^T$ と表し、 \mathbf{x}_i 所与のもとでのハザード関数を $\lambda_i(t|\mathbf{x}_i)$ とおく。このとき、比例ハザードモデルでは、共変量 \mathbf{x}_i をもつ被験者のハザード関数 $\lambda_i(t|\mathbf{x}_i)$ に対し

$$\lambda_i(t|\mathbf{x}_i) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}_i) \quad (1)$$

を想定する。ここに $\lambda_0(t)$ は基準ハザード関数であり、 $\boldsymbol{\beta}=(\beta_1, \beta_2, \dots, \beta_p)^T$ は共変量ベクトル \mathbf{x}_i に対する回帰係数ベクトルである。また、共変量ベクトル \mathbf{x}_i をもつ被験者の生存関数 $S(t|\mathbf{x}_i)$ は、 $\Lambda_0(t)$ を累積基準ハザード関数として

$$S(t|\mathbf{x}_i) = \exp(-\exp(\boldsymbol{\beta}^T \mathbf{x}_i) \Lambda_0(t)) \quad (2)$$

となる。

比例ハザードモデルに対するベイズ流アプローチとしては、回帰係数に対して事前分布を与え部分尤度を用いて推論を行う方法と、回帰係数のみならずハザード関数にも事前情報を与えて推論を行う方法とがある。前者の部分尤度を用いる方法については、黒田(2010)がその数理背景と PHREG プロシジャによる SAS プロ

グラムコードを紹介している．以下では，後者の方法として，時間軸を小さな区間に分け，区間ごとにハザードをモデル化することを考える．時間軸を， $0 = h_0 < h_1 < h_2 < \dots < h_{J-1} < h_J$ と J 個の区間 $[0, h_1), [h_1, h_2), \dots, [h_{J-1}, h_J)$ に分割し， $t \in [h_{j-1}, h_j)$ のとき $\lambda_0(t) = \lambda_j$ ($j=1, 2, \dots, J$) であると考え．これは，時間軸を小さな区間に分割し，各区間内では基準ハザード関数が一定とみなすことを意味する．このとき，モデルパラメータは回帰係数 $\boldsymbol{\beta}$ と基準ハザード関数におけるパラメータ $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_J)^T$ であり，累積基準ハザード関数 $\Lambda_0(t)$ は

$$\Lambda_0(t) = \sum_{j=1}^J \lambda_j \Delta_j(t),$$

ここに

$$\Delta_j(t) = \begin{cases} 0 & (t < h_j) \\ t - h_j & (h_{j-1} \leq t < h_j) \\ h_j - h_{j-1} & (t \geq h_j) \end{cases} \quad (3)$$

である． n 人のデータ $D = \{t_i, v_i, \mathbf{x}_i\}_{i=1}^n$ が与えられたもとの尤度は，ハザード関数 $\lambda_j(t | \mathbf{x}_i)$ ，生存関数 $S(t | \mathbf{x}_i)$ に対応する確率密度関数を $f_i(t | \mathbf{x}_i)$ として，

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda} | D) = \prod_{i=1}^n f_i(t_i | \mathbf{x}_i)^{v_i} S(t_i | \mathbf{x}_i)^{1-v_i} = \prod_{i=1}^n (\lambda_i(t_i | \mathbf{x}_i) S(t_i | \mathbf{x}_i))^{v_i} S(t_i | \mathbf{x}_i)^{1-v_i} = \prod_{i=1}^n \lambda_i(t_i | \mathbf{x}_i)^{v_i} S(t_i | \mathbf{x}_i)$$

と表すことができる．これに(1)から(3)まで代入し整理すれば

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda} | D) = \prod_{i=1}^n \prod_{j=1}^J (\lambda_j \exp(\boldsymbol{\beta}^T \mathbf{x}_i))^{\delta_{ij} v_i} \exp\left[-\delta_{ij} \{\lambda_j(t_i - h_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(t_g - h_{g-1})\}\right] \quad (4)$$

となる(Ibrahim et al., 2001)．ここに δ_{ij} は， i 番目の被験者が j 番目の区間 $(h_{j-1}, h_j]$ にてイベント発現あるいは打ち切りであれば 1，それ以外は 0 をとる関数である ($i = 1, 2, \dots, n; j = 1, 2, \dots, J$)．(4)の形式で表現されるセミパラメトリックモデルは，区分指数モデルとよばれる(Ibrahim et al., 2001)． $\boldsymbol{\beta}$ の事前分布として平均ベクトル $\boldsymbol{\mu}_0$ ，分散共分散行列 $\boldsymbol{\Sigma}_0$ の多変量正規分布を， j 番目の区間における基準ハザード関数 λ_j の事前分布として互いに独立に平均 a_j / b_j ，分散 a_j / b_j^2 のガンマ分布（以下 $Ga(a_j, b_j)$ と表記する）を想定すると， $\boldsymbol{\beta}$ と $\boldsymbol{\lambda}$ の同時事後分布 $p(\boldsymbol{\beta}, \boldsymbol{\lambda} | D)$ は

$$p(\boldsymbol{\beta}, \boldsymbol{\lambda} | D) \propto L(\boldsymbol{\beta}, \boldsymbol{\lambda} | D) p(\boldsymbol{\beta}) p(\boldsymbol{\lambda})$$

で与えられる．ここに $p(\boldsymbol{\beta})$ は $\boldsymbol{\beta}$ の事前分布

$$p(\boldsymbol{\beta}) \propto \exp\left(-(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_0) / 2\right),$$

であり， $p(\boldsymbol{\lambda})$ は $\boldsymbol{\lambda}$ の事前分布

$$p(\boldsymbol{\lambda}) \propto \prod_{j=1}^J \lambda_j^{a_j - 1} \exp(-b_j \lambda_j)$$

である．

PHREG プロシジャでは，BAYES ステートメントによりベイズ流生存時間解析を行うことができる．SAS プログラム 1 は，Gehan (1965) の急性白血病データ（以下「Gehan のデータ」とよぶ）に対し区間指数モデルをあてはめ，ベイズ流生存時間解析を行うための SAS プログラム例である．大橋ら(2016)に倣い，Gehan のデータから，再発または打ち切りを表す変数 **Week**，群を表す変数 **Drug**（プラセボ群であれば 0，6-MP 群であれば 1），打ち切りを表す変数 **Remis**（打ち切りであれば 0，イベント発生であれば 1）をもつデータセット **Gehan** を用意した上での実行を想定している．

SAS プログラム 1 : 区分指数モデルに対するベイズ流生存時間解析

```
data Covar ; do drug = 0,1 ; output ; end ; format drug drugf. ;run ;
proc phreg data = Gehan ;
  class drug ;
  model week*remiss(0) = drug ;
  hazardratio drug ;
  bayes seed = 1234 outpost = outsample nbi = 1000 nmc = 10000 thin = 2
  coeffprior = normal(var = 1e+3) statistics = (summary interval)
  piecewise = hazard(ninterval = 5 prior = gamma(shape = 1e-2 iscale = 1e-2)) ;
  baseline covariates = covar out = out2 survival = Survival
  / nomean group = drug ;
  format drug drugf. ;
run ;
```

BAYES ステートメントにて、乱数のシードを 1234、事後分布からのモンテカルロ標本を保存する SAS データセット名を `Outsample`、burn-in 数を 1,000、生成させるモンテカルロ標本数を 10,000、モンテカルロ標本のサンプリング間隔を 2 と指定した。従って、5,000 個のモンテカルロ標本が `Outsample` に出力される。`coeffprior = normal(var = 1e+3)` と指定したことにより、回帰係数に対する事前分布として、平均 0、分散 1,000 の正規分布を与えたことになる。オプション `PIECEWISE =` が区間指数モデルをあてはめるためのキーワードである。`piecewise = hazard` と指定したことにより、各区間の基準ハザード関数に対して事前分布があてはめられる。`PIECEWISE` オプションにおいて、`prior = gamma(shape = 1e-2, iscale = 1e-2)` と指定したことにより、各区間の基準ハザード関数の事前分布として、平均 1、分散 100 のガンマ分布が用いられる。事後分布からのサンプリングアルゴリズムとして、適応的棄却メトロポリスサンプリング (adaptive rejection Metropolis sampling ; ARMS)、ランダムウォークメトロポリスアルゴリズムのいずれかが使用可能であり、特に指定しない限り ARMS が適用される。ARMS は、Gilks et.al. (1995) により提案されたサンプリングアルゴリズムであり、メトロポリス - ヘイスティングアルゴリズムと適応的棄却サンプリング (Adaptive rejection sampling ; ARS) とを組み合わせた方法である。

出力結果 1 及び出力結果 2 は、SAS プログラム 1 により得られる出力結果から一部抜粋したものである。オプション `PIECEWISE =` において `ninterval = 5` と指定したため、時間軸が 5 個の区間に分割され、基準ハザード関数は 5 個のパラメータによりモデル化される。出力結果 1 より、時間軸を [0,4.5), [4.5,6.5), [6.5,10.5), [10.5,15.5), [15.5,∞) の 5 つの区間に分割し、各区間の基準ハザード関数 λ_j ($j=1,2,\dots,5$) に対して事後更新が行われたことが確認できる。区間数の代わりに分割したい区間を直接指定したい場合には、`intervals =` とすればよい。BAYES ステートメントにて `statistics = (summary interval)` と指定したことにより、事後分布からの要約統計量として、モデルパラメータごとに、モンテカルロ標本数、平均、標準偏差、パーセント点、95% 信用区間、95%HPD (HPD : highest posterior density) 信用区間が算出される (出力結果 2)。SAS プログラム 1 のように、BAYES ステートメントと HAZARDRATIO ステートメントとを同時に指定することで、ハザード比に関するベイズ流の推定値を出力することが可能である。OUTPUT 画面にて「ハザード比 drug」とタイト

ルされた出力結果から、対照群に対する 6-MP 群のハザード比の事後平均 0.1985, 95%信用区間(0.0764, 0.4049), 95%HPD 信用区間(0.0552, 0.3707)となることがわかる。

出力結果 1：区間指数モデルをあてはめたときの時間区切り

[下限,	上限)	N	事象	ハザードパラメータ
0	4.5	7	7	Lambda1
4.5	6.5	6	5	Lambda2
6.5	10.5	8	6	Lambda3
10.5	15.5	7	6	Lambda4
15.5	Infy	14	6	Lambda5

出力結果 2：区分指数モデルに対するベイズ流生存時間解析 事後分布に関する要約統計量

パラメータ	平均	標準偏差	パーセント点			信用区間			
			25%	50%	75%	95% 信用区間	95% HPD		
Lambda1	0.0712	0.0277	0.0514	0.0674	0.0870	0.0290	0.1366	0.0234	0.1256
Lambda2	0.1545	0.0718	0.1029	0.1439	0.1940	0.0482	0.3283	0.0388	0.2983
Lambda3	0.1209	0.0516	0.0841	0.1134	0.1502	0.0426	0.2408	0.0342	0.2217
Lambda4	0.1737	0.0745	0.1202	0.1630	0.2169	0.0594	0.3471	0.0454	0.3201
Lambda5	0.1752	0.0833	0.115	0.1619	0.2205	0.0556	0.3741	0.0457	0.3495
drug6_MP	-1.7075	0.4307	-1.9982	-1.6939	-1.4078	-2.5724	-0.9042	-2.5516	-0.8901

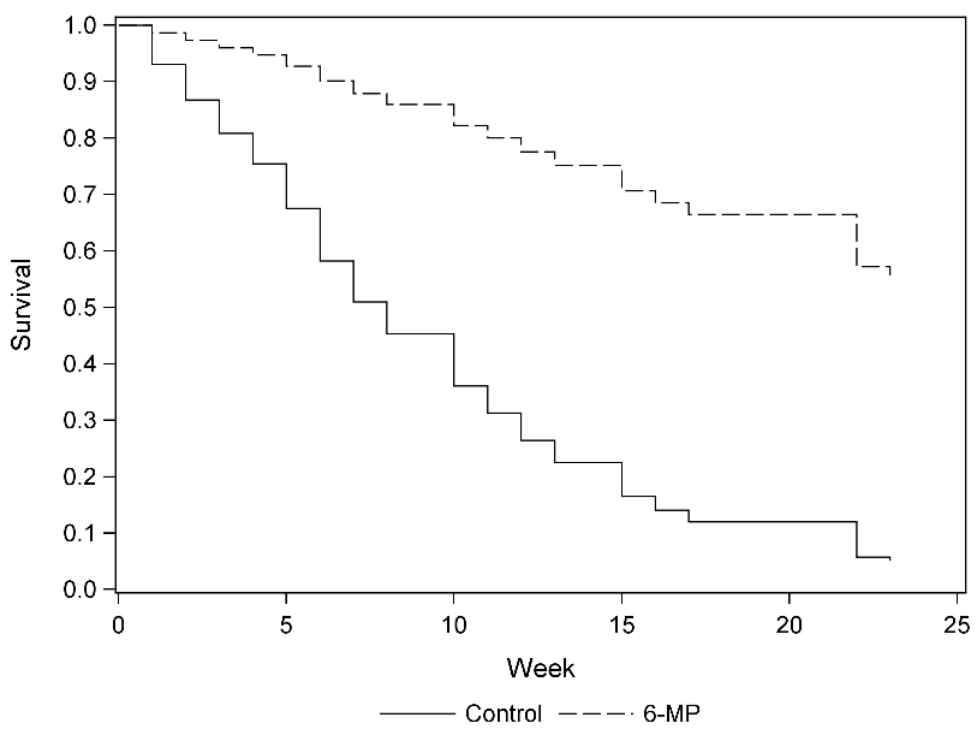


図 1：BASELINE ステートメントによる生存時間曲線

図1は、BASELINE ステートメントにより出力された予測生存時間曲線をプロットしたものである。データセット Out2 を用いて SGPLOT プロシジャにより作成した。BASELINE ステートメントにおいて、DIRADJ オプションを指定することで、GROUP = で指定した変数別に直接調整された生存時間曲線が可能である(大橋ら, 2016; 魚住ら, 2017)。しかし、バイズ流生存時間解析を行う際には DIRADJ オプションが無効となるため、COVARIATES = では、別途用意した群を表す共変量の値をもたせたデータセット Covar を指定している。

3. フレイルティモデル

フレイルティモデルは、比例ハザードモデルに変量効果をもたせたモデルである。例えば、眼に起こるある疾患をイベントとする生存時間データを解析したいとする。眼を個体とすれば、個体に生起するイベントは1つとなる。患者をクラスターと捉え、右眼と左眼でのイベントのおこりやすさは、クラスターの異質性に起因すると考える。クラスターの異質性を変量効果とみなし(生存時間解析ではフレイルティとよばれる)、比例ハザードモデルに反映させる。

フレイルティモデルは、以下のように定式化される。

i 番目のクラスターにおける j 番目の個体に対して、 p 個の共変量を $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp})^T$ 、ハザード関数を $\lambda_{ij}(t|\mathbf{x}_{ij})$ とする ($i = 1, 2, \dots, G; j = 1, 2, \dots, n_i$)。 i 番目のクラスターのフレイルティを w_i とし、基準ハザード関数 $\lambda_0(t)$ を用いて

$$\lambda_{ij}(t|\mathbf{x}_{ij}) = \lambda_0(t) w_i \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta}) \quad (5)$$

とモデル化する。(5)において w_i が互いに独立に平均 1, 分散 θ のガンマ分布 $Ga(1/\theta, 1/\theta)$ を想定したモデルは、ガンマフレイルティモデルとよばれる。また、(5)において $\gamma_i = \log w_i$ とおけば

$$\lambda_{ij}(t|\mathbf{x}_{ij}) = \lambda_0(t) \exp(\gamma_i + \mathbf{x}_{ij}^T \boldsymbol{\beta}) \quad (6)$$

となる。(6)において γ_i が互いに独立に $N(0, \sigma^2)$ に従うと想定したモデルを、対数正規フレイルティモデルとよぶ。 $\boldsymbol{\beta}$ の事前分布を $p(\boldsymbol{\beta})$ 、 $\mathbf{w} = (w_1, w_2, \dots, w_G)^T$ の分布を規定するパラメータ θ の事前分布を $p(\theta)$ 、パラメータ θ 所与のもとでの \mathbf{w} の同時確率密度関数を $g(\mathbf{w} | \theta)$ 、得られたデータ D に対する部分尤度を $L_p(\boldsymbol{\beta}, \mathbf{w} | D)$ と表すとき、モデルパラメータの事後分布 $p(\boldsymbol{\beta}, \mathbf{w}, \theta | D)$ は

$$p(\boldsymbol{\beta}, \mathbf{w}, \theta | D) \propto L_p(\boldsymbol{\beta}, \mathbf{w} | D) g(\mathbf{w} | \theta) p(\boldsymbol{\beta}) p(\theta) \quad (7)$$

で与えられる。

フレイルティモデルの適用例として、Lin (1994)の糖尿病性網膜症のデータを取りあげる。このデータ(以下「糖尿病性網膜症のデータ」とよぶ)は、糖尿病性網膜症患者に対するレーザー治療の効果を比較するため、失明までの時間を評価項目として、National Eye Institute で行われた試験の 197 名のデータであり、大橋ら (2016) の第 3 章や SAS/STAT(R) User's Guide, The PHREG Procedure においても用いられている。患者にはどちらかの眼にランダムにレーザー治療が割付けられ、もう片方の眼はレーザー治療なしの状態、失明までの時間が観察された。本稿では、失明または打ち切りまでの時間を表す変数 Time, 打ち切りを表す変数 Status (打ち切りであれば 0, イベント発生であれば 1), レーザー治療あり・なしを表す変数 Treat (レーザー治療なしのとき 0, レーザー治療ありのとき 1), 基礎疾患の発症時期を表す変数 Type (発症時期が 20 歳未満なら 0, 20 歳以上なら 1)をもつ SAS データセット Diab を用いる。

SAS プログラム 2 : フレイルティモデルに対するベイズ流生存時間解析

```

proc phreg data = Diab ;
  class ID Treat Type ;
  model Time*Status(0) = Treat Type Treat*Type ;
  hazardratio Treat ;
  random ID / dist = lognormal ;
  bayes seed = 1234 outpost = outsample nbi = 1000 nmc = 10000 thin = 1
  coeffprior = normal(var = 1e+6) statistics = (summary interval)
  dispersionprior = igamma(shape = 1e-2, scale = 1e-2) statistics = (summary interval) ;
  format Type Typef. Treat Lazerf. ;
run ;

```

SAS プログラム 2 は、糖尿病性網膜症のデータに対し、治療法、基礎疾患の発症時期、治療法と基礎疾患の発症時期の交互作用を共変量とし、クラスター（被験者）のフレイルティを含めた、下記対数正規フレイルティモデルによるベイズ流生存時間解析を行うためのプログラム例である。

$$\lambda_{ij}(t | x_{ij1}, x_{ij2}) = \lambda_0(t) \exp(\gamma_i + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij1}x_{ij2}) \quad (i=1,2,\dots,197; j=1,2)$$

$$\beta_1, \beta_2, \beta_3 \sim N(0, 10^6)$$

$$\gamma_i | \theta \sim N(0, \theta)$$

$$\theta \sim IG(0.01, 0.01)$$

ここに、 x_{ij1}, x_{ij2} は、 i 番目のクラスター（患者）の j 番目の個体（右眼あるいは左眼）における治療法、基礎疾患の発症時期を表し、 γ_i が i 番目のクラスター（患者）のフレイルティである。

RANDOM ステートメントにてクラスターを表す変数を指定し、オプション DIST = により分布を指定することで、クラスターがある特定の分布に従うものとして、モデルに反映することが可能となる。SAS プログラム 2 では、デフォルトである dist = lognormal と指定しており、(6)によって表される対数正規フレイルティモデルが設定される。なお、dist = gamma と指定すると、(5)によって表されるガンマフレイルティモデルが設定される。BAYES ステートメントにおいて、パラメータ θ の事前分布 $p(\theta)$ を指定するオプションが DISPERSIONPRIOR = である。gamma(scale = a, iscale = b) とすれば $Gam(a, b)$ を、igamma(scale = a, scale = b) とすれば平均 $b/(a-1)$ 、分散 $b^2/((a-1)^2(a-2))$ の逆ガンマ分布（以下 $IG(a, b)$ と表記する）を、それぞれ指定することができる。PHREG プロシジャでフレイルティモデルに対するベイズ流生存時間解析を行うと、事後分布からのサンプリングには、ランダムウォークメトロポリスアルゴリズムが適用される。

出力結果 3 : フレイルティモデルに対するベイズ流生存時間解析 事後分布に関する要約統計量

パラメータ	平均	標準偏差	パーセント点			信用区間			
			25%	50%	75%	95% 信用区間	95% HPD		
β_1	-0.5056	0.2329	-0.6637	-0.5137	-0.3444	-0.9612	-0.0571	-0.9238	-0.0332
β_2	0.4437	0.2725	0.2786	0.4409	0.6145	-0.1027	0.9984	-0.1125	0.9708
β_3	-1.0518	0.3896	-1.3135	-1.0637	-0.8009	-1.7819	-0.2413	-1.7892	-0.2686
θ	1.1576	0.4267	0.8682	1.1247	1.4134	0.2398	2.0801	0.1642	1.9070

出力結果 3 は、モデルパラメータの事後分布に関する要約統計量である。事後確率 $\Pr(\beta_l > 0|D)$ ($l=1,2,3$)は、事後分布から生成された R 個のモンテカルロ標本を $\beta_l^{(1)}, \beta_l^{(2)}, \dots, \beta_l^{(R)}$ として

$$\Pr(\beta_l > 0|D) = \frac{1}{R} \sum_{r=1}^R I(\beta_l^{(r)} > 0)$$

により求めることができる(Albert, 2009)。ここで、 $I(\cdot)$ はカッコの中が成り立てば 1 を、そうでなければ 0 をとる指示関数である。SAS プログラム 2 によって得られる、事後分布からのモンテカルロ標本を出力したデータセット `Outsample` を用いて算出すると、 $\Pr(\beta_1 > 0|D) = 0.0069$, $\Pr(\beta_2 > 0|D) = 0.9408$, $\Pr(\beta_3 > 0|D) = 0.0004$ となった。OUTPUT 画面にて「Treat のハザード比」とタイトルされた出力結果から、レーザー治療なしに対するレーザー治療ありのハザード比は、20 歳以降に発症した場合、事後平均 0.2211, 95%信用区間(0.1163, 0.3893), 95%HPD 信用区間(0.1101, 0.3655)であるのに対し、20 歳未満で発症した場合、事後平均 0.6198, 95%信用区間(0.3824, 0.9445), 95%HPD 信用区間(0.3658, 0.9066)であった。

上記は、ハザード関数には事前情報を付与せず、回帰係数にのみ事前情報を与えて、部分尤度に基づいて推論を行う方法である。比例ハザードモデルに対するアプローチと同様に、ハザード関数及び回帰係数に事前情報を付与し、これらの同時確率分布を用いて推論を行う方法がある。以降では、その一例として、ハザード関数に事前情報を付与するため、時間軸を小さな区間に分け、区間ごとにハザードをモデル化した、区分指数フレイルティモデル (piecewise exponential frailty model) について紹介する。

i 番目のクラスターにおける j 番目の個体に対して、イベントが観察されるまでの時間 t_{ij} , 観察打ち切りまでの時間 c_{ij} に対し、 $u_{ij} = \min(t_{ij}, c_{ij})$, $\omega_{ij} = I(t_{ij} \leq c_{ij})$ として、 $N_{ij}(t) = I(t_{ij} \leq t, \omega_{ij}=1)$, $Y_{ij}(t) = I(t_{ij} \geq t)$ とおく。ここに $I(\cdot)$ はカッコの中が成り立てば 1 を、そうでなければ 0 をとる指示関数である。フレイルティモデルにおけるハザード関数 $\lambda_{ij}(t|\mathbf{x}_{ij})$ を、強度関数 $I_{ij}(t)$ で置きかえると

$$I_{ij}(t) = Y_{ij}(t) \lambda_0(t) w_i \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta})$$

となる。打ち切りが non-informative であるとき、生存時間が互いに独立として尤度は

$$\prod_{i=1}^n \prod_{j=1}^{n_i} \left[\prod_{t>0} I_{ij}(t) \right] e^{-\int_0^{c_{ij}} I_{ij}(t) dt} \quad (8)$$

となる。ここで、 $dN_{ij}(t)$ は微小時間 $[t, t+dt)$ に対する N_{ij} の増分であり、 i 番目のクラスターにおける j 番目の個体が dt の間にイベントが生起すれば 1, そうでなければ 0 の、いずれかの値をとる。累積基準ハザード関数 $\Lambda_0(t)$ の微小時間 $[t, t+dt)$ に対する増分を $d\Lambda_0(t)$ と表すとき、微小時間 $[t, t+dt)$ に対する強度関数の増分 $I_{ij}(t)dt$ は

$$I_{ij}(t)dt = Y_{ij}(t) w_i \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta}) d\Lambda_0(t) \quad (9)$$

となり、尤度(8)は、(9)で表される $I_{ij}(t)dt$ を平均値とするポアソン分布の核 $[I_{ij}(t)dt]^{dN_{ij}(t)} \exp(-I_{ij}(t)dt)$ から構成されることがわかる。

時間軸を、 $0 = h_0 < h_1 < h_2 < \dots < h_{K-1} < h_K$ と K 個の区間 $[0, h_1), [h_1, h_2), \dots, [h_{K-1}, h_K)$ に分割し、 $t \in [h_{k-1}, h_k)$ のとき $\lambda_0(t) = \lambda_k$ ($k=1,2,\dots,K$) であると想定する。(8), (9)の微小時間 $[t, t+dt)$ を区間 $[h_{k-1}, h_k)$ と置き換えることで、(8)により表される尤度を算出することができる。フレイルティモデルの場合、 $\boldsymbol{\beta}$ の事前分布、基準ハザード $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_K)^T$ の事前分布のほか、 i 番目のクラスター ($i=1,2,\dots,G$) のフレイルティにある特定の分布を想定し、その分布を規定するパラメータ θ の事前分布を与えることになる。

現時点では、PHREG プロシジャで区分指数フレイルティモデルを実行することはできない。ただし、DATA ステップを用いて入力データを加工することで、MCMC プロシジャによる実行が可能である。SAS/STAT(R) User's Guide, The MCMC Procedure では、糖尿病性網膜症のデータに対し、区分指数フレイルティモデル(10)

をあてはめてバイズ流生存時間解析を行うためのプログラムコード (SAS プログラム 3) が紹介されている。

$$\begin{aligned} \lambda_{ij}(t | x_{ij1}, x_{ij2}) &= \lambda_0(t) \exp(\gamma_i + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij1}x_{ij2}) \quad (i=1,2,\dots,197; j=1,2) \\ \lambda_k &\sim \text{Gam}(0.01, 0.01) \quad (k=1,2,\dots,8) \\ \beta_1, \beta_2, \beta_3 &\sim N(0, 10^6) \\ \gamma_i | \theta &\sim N(0, \theta) \\ \theta &\sim \text{IG}(0.01, 0.01) \end{aligned} \tag{10}$$

SAS プログラム 3 では、時間軸を $0.1 < 6.545 < 13.95 < 26.47 < 38.8 < 45.88 < 54.35 < 62 < 80$ と、8つの区間 $[0.1, 6.545)$, $[6.545, 13.95)$, $[13.95, 26.47)$, $[26.47, 38.8)$, $[38.8, 45.88)$, $[45.88, 54.35)$, $[54.35, 62)$, $[62, 80)$ に分割している。分割値 $(0.1, 6.545, 13.95, 26.47, 38.8, 45.88, 54.35, 62, 80)$ をもたせたデータセット **Partition** を用意する。データセット **Partition** と **Diab** から、197 個のクラスター (被験者) における 2 つの個体 (右眼と左眼)、合計 $197 \times 2 = 394$ 個のデータに対し、各分割値でのリスクセットを変数 **Y** に、計数過程を変数 **dN** に、区間を識別する番号を変数 **INT_INDEX** に、それぞれ入力したデータセット **Inputdata** を用意する。その上で、**MCMC** プロシジャを用いてモデルパラメータの推論を行う。**MCMC** プロシジャでは、**RANDOM** ステートメントにより、クラスターのフレイルティ、基準ハザードに対して想定する分布を指定し、クラスターごと、区間ごとに、それぞれ互いに独立に分布をあてはめることを、オプション **SUBJECT =** を用いて設定している。

SAS プログラム 3 : 区分指数フレイルティモデルに対するバイズ流生存時間解析

```
data partition ;
input int_1-int_9 ;
datalines;
0.1 6.545 13.95 26.47 38.8 45.88 54.35 62 80
;
run ;
data _a ;set Diab ;
  if _n_ eq 1 then set partition ;
  array int[*] int_1-int_9 ; array Y[8] ; array dN[8] ;
  do k = 1 to (dim(int)-1) ;
    Y[k] = (time - int[k] + 0.001 >= 0) ; dN[k] = Y[k] * ( int[k+1] - time - 0.001 >= 0) * status ;
  end ;
  output ;
  drop int_1-int_9 ;
run ;
data _b ;set _a ;
  array y[*] y_1-y_9 ; array dn[*] dn_1-dn_9 ;
  do i = 1 to (dim(y)) ; y_val = y[i] ; dn_val = dn[i] ; int_index = i ;output ;end ;
  keep y_1-y_9 dn_1-dn_9 type treat int_index id ;
run ;
data _b ;set _b ;rename y_val = Y dn_val = dN ;run ;
```

```

data inputdata ;set _b ;if Y > 0 ;run ;
proc mcmc data = inputdata nmc = 10000 outpost = postout seed = 12351 stats = summary diag = none ;
  parms beta1-beta3 0 s2 ;
  prior beta: ~ normal(0, var = 1e+6) ;
  prior s2 ~ igamma(0.01, scale = 0.01) ;
  random lambda ~ gamma(0.01, iscale = 0.01) subject = int_index ;
  random u ~ normal(0, var = s2) subject = id ;
  bZ = beta1*treat + beta2*type + beta3*treat*type + u ;
  idt = exp(bZ) * lambda ;
  model dN ~ poisson(idt) ;
run ;

```

出力結果 4：区分指数フレイルティモデルに対するベイズ流生存時間解析 事後分布に関する要約統計量

パラメータ	平均	標準偏差	パーセント点			信用区間			
			25%	50%	75%	95% 信用区間	95% HPD		
β_1	-0.4341	0.2290	-0.5882	-0.4274	-0.2873	-0.9034	0.0239	-0.9658	-0.0453
β_2	0.3552	0.2194	0.2194	0.2069	0.5000	-0.0529	0.8053	-0.0536	0.8022
β_3	-0.3529	0.3477	-1.0927	-1.0927	-0.6136	-1.5162	-0.1793	-1.5600	-0.2312
θ	0.3434	0.2063	0.1999	0.3137	0.4547	0.0422	0.8143	0.0267	0.7158

出力結果 4 は、SAS プログラム 3 を実行して得られる結果をまとめたものである (SAS プログラム 3 で、stats = summary を stats = all とすれば信用区間が得られる)。SAS プログラム 3 では、ハザード比は算出されないため、モンテカルロ標本から算出することになる。(10)では、レーザー治療なしのとき $x_{ij1} = 0$ 、レーザー治療なしのとき $x_{ij1} = 1$ 、発症時期 20 歳未満のとき $x_{ij2} = 0$ 、発症時期 20 歳以上のとき $x_{ij2} = 1$ とコーディングしているので、20 歳以降に発症した場合でのレーザー治療なしに対するレーザー治療ありのハザード比 HR_1 は

$$\frac{\exp(\beta_1 \times 1 + \beta_2 \times 1 + \beta_3 \times 1 \times 1)}{\exp(\beta_1 \times 0 + \beta_2 \times 1 + \beta_3 \times 0 \times 1)} = \exp(\beta_1 + \beta_3)$$

となる。従って、事後分布からの R 個のモンテカルロ標本を $\beta_1^{(1)}, \beta_1^{(2)}, \dots, \beta_1^{(R)}$ として ($l = 1, 2, 3$)、ハザード比 HR_1 の事後平均は、 $\exp(\beta_1^{(1)} + \beta_3^{(1)})$, $\exp(\beta_1^{(2)} + \beta_3^{(2)})$, ..., $\exp(\beta_1^{(R)} + \beta_3^{(R)})$ の平均値として求めることができ、95%信用区間はこれら R 個のモンテカルロ標本の 2.5%点と 97.5%点として求めることができる。一方、20 歳未満で発症した場合でのレーザー治療なしに対するレーザー治療ありのハザード比 HR_2 は

$$\frac{\exp(\beta_1 \times 1 + \beta_2 \times 0 + \beta_3 \times 1 \times 0)}{\exp(\beta_1 \times 0 + \beta_2 \times 0 + \beta_3 \times 0 \times 0)} = \exp(\beta_1)$$

となる。よってハザード比 HR_2 の事後平均は $\exp(\beta_1^{(1)})$, $\exp(\beta_1^{(2)})$, ..., $\exp(\beta_1^{(R)})$ の平均値として、95%信用区間はこれら R 個のモンテカルロ標本の 2.5%点と 97.5%点として、それぞれ求めることができる。SAS プログラム 3 により作成されるデータセット Outpost を用いて算出すると、20 歳以降に発症した場合でのレーザー治療なしに対するレーザー治療ありのハザード比 HR_1 の事後平均 0.2866、95%信用区間(0.1587,0.4645)、20 歳未満に発症した場合でのレーザー治療なしに対するレーザー治療ありのハザード比 HR_2 の事後平均 0.6650、95%

信用区間(0.4052, 1.0242)であった.

4. まとめ

生存時間解析の手法として, 比例ハザードモデルとフレイルティモデルをとりあげ, ベイズ流アプローチについて紹介した. 比例ハザードモデルについては, 時間軸をいくつかの区間に分割することでハザードを複数のパラメータで表現する, 区分指数モデルをとりあげた. 区分指数モデルは, PHREG プロシジャの BAYES ステートメントにおいて, PIECEWISE オプションを用いることで実行可能である. なお, 区分指数モデルはパラメトリックに例数設計を行う際にも有用である (魚住ら, 2009). フレイルティモデルに対して, 部分尤度に基づく推論を行いたい場合には, PHREG プロシジャの BAYES ステートメントにおいて, COEFFPRIOR オプションと DISPERSIONPRIOR オプションを用いることで実行可能である. 一方, フレイルティモデルにおいて, ハザード関数にも事前情報を与えて解析を行うというアプローチも考えられる. その一例として, 区分指数フレイルティモデルをとりあげた. 区分指数フレイルティモデルは, 計数過程に対応させた形式に加工した入力データを用いて, MCMC プロシジャの RANDOM ステートメントにより実行することができる.

ベイズ流生存時間解析には, これ以外にも多数の手法が存在する. 本稿で説明した区分指数モデルでは, 各区間での基準ハザード関数は互いに独立と仮定している. しかし, 隣接する区間でのハザードは同じような傾向をもつと想定し, 基準ハザード関数に対して相関をもつ事前分布を適用することも考えられる. また, 累積ハザードに対して事前情報を付与する方法も考えられる. ガンマ過程 (Kalbfleisch, 1978; Clayton, 1991), 拡張ガンマ過程 (Dykstra and Laud, 1981), ベータ過程 (Hjort, 1990) といった確率過程を用いた累積ハザード関数のモデル化が該当する. 事前確率過程を用いる場合, 確率密度関数を明示的に導出できないため, 推論はシミュレーション実験に基づくものとなる. このため, 各モデルに応じた事後確率過程からのサンプリングアルゴリズムが開発・研究されている. ガンマ過程を用いたベイズ流生存時間解析は, MCMC プロシジャを用いて実装が可能である (矢田・浜田, 2016).

本稿では, 生存時間データに対するベイズ流解析のうち, PHREG プロシジャや MCMC プロシジャの既存の機能で実行可能な手法について説明した. PHREG プロシジャや MCMC プロシジャは, SAS/STAT として提供されている. SAS/STAT は毎年のようにバージョンアップされ, そのたびに新規プロシジャの追加, 既存プロシジャでの新機能搭載がなされている (大橋ら, 2016; 浜田・魚住, 2016; 魚住ら, 2016). 今後, SAS/STAT の機能拡張が行われていく中で, ベイズ流解析の手法も組み込まれ, ベイズ流アプローチがより身近になることを期待したい.

参考文献

- [1]大橋靖雄・浜田知久馬 (1995). 生存時間解析 –SAS による生物統計. 東京大学出版会.
- [2]大橋靖雄・浜田知久馬・魚住龍史 (2016). 生存時間解析 応用編 –SAS による生物統計. 東京大学出版会.
- [3]Cox, D.R. (1972). Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society, Series B*. **34**:187–220.
- [4]SAS Institute Inc. (2008). *SAS/STAT(R) 9.2 User's Guide*, Cary, NC, USA: SAS Institute Inc.

- [5]矢田真城・魚住龍史・浜田知久馬 (2016). 生存時間データに対するベイズ流例数設計. SAS ユーザー総会 論文集, 268-278.
- [6]SAS Institute Inc. (2011). *SAS/STAT(R) 9.3 User's Guide*, Cary, NC, USA: SAS Institute Inc.
- [7]黒田普吾 (2010). Cox 回帰における Bayes 推定と PHREG プロシジャ. SAS ユーザー総会 論文集, 23-44.
- [8]SAS Institute Inc. (2016). *SAS/STAT(R) 14.2 User's Guide*, Cary, NC, USA: SAS Institute Inc.
- [9]Ibrahim, J.G., Chen, M.H., Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag, New York.
- [10]Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrary single-censored samples. *Biometrika*. **52**: 203-223.
- [11]Gilks, W. R., N. G. Best, and K. K. C. Tan. (1995). Adaptive Rejection Metropolis Sampling with Gibbs Sampling. *Journal of the Royal Statistical Society*, **44**: 455-472.
- [12]魚住龍史・矢田真城・浜田知久馬 (2017). PHREG プロシジャにおける共変量調整解析に関連したオプション機能. SAS ユーザー総会 論文集.
- [13]Lin, D. Y. (1994). Cox Regression Analysis of Multivariate Failure Time Data: The Marginal Approach. *Statistics in Medicine*, **13**: 2233-2247.
- [14]Albert, J. Bayesian Computation with R. Springer-Verlag, 2009. 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, 2012.
- [15]魚住龍史・水澤純基・浜田知久馬 (2009). 生存時間解析における Lakatos の症例数設計法の有用性の評価. SAS ユーザー総会 論文集, 143-152.
- [16]Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society: Series B*. **40**:214-221.
- [17]Clayton, D. G. (1991). A Monte Carlo Method for Bayesian Inference in Frailty Models. *Biometrics*, **47**:467-485.
- [18]Dykstra, R.L. and Laud, P.W. (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics*. **9**, 356-367.
- [19]Hjort, N.L. (1990). Nonparametric Bayes estimations based on beta processes in models for life history data. *The Annals of Statistics*. **18**, 1259-1294.
- [20]矢田真城・浜田知久馬 (2016). ガンマ過程を用いた比例ハザードモデルによる解析. SAS ユーザー総会 論文集, 100-109.
- [21]浜田知久馬・魚住龍史 (2016). SAS による生存時間分布の予測「Death Note の統計学」. SAS ユーザー総会論文集, 297.
- [22]魚住龍史・矢田真城・浜田知久馬 (2016). SAS プロシジャを用いた生存時間データに対する例数設計の变革. SAS ユーザー総会 論文集, 250-267.

連絡先

E-mail: yada-s@a2healthcare.com