

ガンマ過程を用いた比例ハザードモデルによる解析

○矢田 真城¹, 浜田 知久馬²

(¹エイツーヘルスケア株式会社, ²東京理科大学)

Proportional hazard model using gamma process

Shinjo Yada¹ and Chikuma Hamada²

¹A2 Healthcare Corporation, ²Tokyo University of Science

要旨

比例ハザードモデルは、生存時間解析において広く用いられている統計モデルである。生存時間分布（基準ハザード関数）に特定の分布を仮定することなく、回帰パラメータの推論を行うことが最大の利点である。比例ハザードモデルに対するベイズ流解析としては、大きく2つのアプローチが存在する。ひとつは、基準ハザード関数に確率モデルを仮定せず、回帰パラメータにのみ確率モデルをあてはめ、部分尤度に基づきパラメータの推論を行うアプローチである。もうひとつは、回帰パラメータのみならず基準ハザード関数にも確率モデルをあてはめ、パラメータの推論を行うアプローチである。前者のアプローチに関しては、PHREGプロシジャにおいてBAYESステートメントを指定することにより実行可能となった。そこで本稿では、後者のアプローチの1つである、累積基準ハザード関数にガンマ過程を想定したモデルをとりあげ、解析方法についてまとめ、SASによる使用例を紹介する。

キーワード：比例ハザードモデル，ガンマ過程，基準ハザード

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）。

生存時間解析において用いられる統計モデルのひとつに、Cox(1972)の比例ハザードモデルがあげられる。比例ハザードモデルには、1)基準ハザード関数が未知のままでも共変量の影響を推測することができる、2)生存時間の順位情報のみに基づいて推測を行うことができるため外れ値に対して頑健である、3)モデルが単純な場合にはログランク検定に一致する（大橋・浜田,1995）、といったいくつかの利点がある。比例ハザードモデルに対しては、基準ハザード関数に特定の分布を仮定することなく、部分尤度に基づき共変量の回帰パラメータを推定する、セミパラメトリックな方法が広く用いられている。その際、基準ハザード関数は局外パラメータとして扱われており、必要に応じて部分尤度法により求められた回帰パラメータの推定量を用

いて基準ハザード関数を推定することができる (中村,2001 ; 大橋・浜田,1995).

一方, 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, 回帰パラメータのみに確率モデルを考えるアプローチと, 回帰パラメータ及び基準ハザード関数に確率モデルを考えるアプローチとがある (Dey and Rao,2005). 前者では, 基準ハザード関数に特定の事前分布をおくことなく, 回帰パラメータのみに事前分布をおき, 部分尤度を用いることで回帰パラメータの事後分布を導出する. SAS では, Ver.9.2 より PHREG プロシジャにおいて BAYES ステートメントを記述することにより, 部分尤度を用いた回帰パラメータの事後分布の算出が可能となった (SAS Institute Inc., 2008 ; 黒田,2010). 本稿では, もうひとつのアプローチである, 回帰パラメータ及び基準ハザード関数に確率モデルをおき, モデルパラメータの推論を行う方法について説明する. 基準ハザード関数に想定する確率過程に関して多数の方法が提案されているが, ここではそれらの先駆けとなった, 累積基準ハザード関数にガンマ過程をあてはめるモデルをとりあげる.

2. 比例ハザードモデルに対するベイズ流解析

2.1. Intensity function による解析

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.

いま i 番目の被験者に対して, イベントが観察されるまでの時間 t_i , 観察打ち切りまでの時間 c_i に対し, $u_i = \min(t_i, c_i)$, $\omega_i = I(t_i \leq c_i)$ として, $N_i(t) = I(t_i \leq t, \omega_i = 1)$, $Y_i(t) = I(t_i \geq t)$ とおく ($i=1,2,\dots,n$). 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. $N_i(t)$ は, i 番目の被験者について時間 t 直前までイベントが生起しなければ 0, イベントが生起しているなら 1 をとる階段関数である. 一方, $Y_i(t)$ は i 番目の被験者について時間 t 直前までイベントが生起していない, つまりイベントが生起するリスクにさらされていたのであれば 1 を, そうでなければ 0 をとる関数である. 微小時間 dt に対する N_i の増分

$$dN_i(t) = N_i(t+dt) - N_i(t)$$

は, i 番目の被験者が微小時間 dt の間にイベントが生起すれば 1, そうでなければ 0 の 2 つの値しかとらない. i 番目の被験者における p 個の共変量を $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{ip})^T$ と表し, \mathbf{z}_i 所与のもとでのハザード関数を $\lambda_i(t | \mathbf{z}_i)$ とおくと, i 番目の被験者に対し時間 t 直前までイベントが生起しておらず ($N_i(t)=0$) リスク下にある状態 ($Y_i(t)=1$) で, 微小時間 dt の間にイベントが生起する ($dN_i(t)=1$) 確率は

$$\Pr(dN_i(t)=1 | Y_i(t)=1, N_i(t)=0, \mathbf{z}_i) = \lambda_i(t | \mathbf{z}_i) dt \quad (2.1)$$

となる. 同様に, i 番目の被験者が打ち切り症例であれば

$$\Pr(dN_i(t)=1 | Y_i(t)=0, N_i(t)=0, \mathbf{z}_i) = 0 \quad (2.2)$$

であり, 時間 t 直前にイベントが生起していれば

$$\Pr(dN_i(t)=1 | Y_i(t)=0, N_i(t)=1, \mathbf{z}_i) = 0 \quad (2.3)$$

である. $N_i(t), Y_i(t)$ はともに 0 か 1 の値をとるため, 組み合わせとしては $\Pr(dN_i(t)=1 | Y_i(t)=1, N_i(t)=1, \mathbf{z}_i)$ も考えられるが, 当該症例において時間 t 直前までにイベントが生起しており, かつリスク下にある状態はありえない. 時間 t までに観測された利用可能なデータ $N_i(t), Y_i(t), \mathbf{z}_i$ をまとめて F_{i-} と表記するとき, (2.1)から(2.3)をまとめて

$$\Pr(dN_i(t)=1 | F_{i-}) = Y_i(t) \lambda_i(t | \mathbf{z}_i) dt$$

と表すことができ,

$$E(dN_i(t)=1 | F_{i-}) = Y_i(t) \lambda_i(t | \mathbf{z}_i) dt \quad (2.4)$$

となる (中村, 2001). i 番目の被験者に対する時間 t での強度関数 (intensity function) を

$$I_i(t)dt = E(dN_i(t) = 1 | F_{t-})$$

と定義するとき, (2.4)より

$$I_i(t)dt = Y_i(t)\lambda_i(t | \mathbf{z}_i)dt$$

よって

$$I_i(t) = Y_i(t)\lambda_i(t | \mathbf{z}_i)$$

となり, ハザード関数 $\lambda_i(t | \mathbf{z}_i)$ を強度関数 $I_i(t)$ に取り替えることができる.

Cox (1972) の比例ハザードモデルでは, 共変量 \mathbf{z}_i をもつ被験者のハザード関数 $\lambda_i(t | \mathbf{z}_i)$ に対し

$$\lambda_i(t | \mathbf{z}_i) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}_i) \quad (2.5)$$

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{z}_i に対する回帰パラメータベクトルである. したがって比例ハザードモデルでは強度関数 $I_i(t)$ は

$$I_i(t)dt = Y_i(t)\lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}_i)dt \quad (2.6)$$

よって

$$I_i(t) = Y_i(t)\lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}_i)$$

となる. また, 共変量ベクトル \mathbf{z}_i をもつ被験者における生存関数 $S(t | \mathbf{z}_i)$ は, 生存関数, ハザード関数, 累積ハザード関数の関係から

$$\begin{aligned} S(t | \mathbf{z}_i) &= \exp(-\Lambda(t | \mathbf{z}_i)) = \exp\left(-\int_0^t \lambda(t | \mathbf{z}_i)dt\right) = \exp\left(-\int_0^t \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{z}_i)dt\right) \\ &= \exp\left(-\exp(\boldsymbol{\beta}^T \mathbf{z}_i) \int_0^t \lambda_0(t)dt\right) = \left[\exp\left(-\int_0^t \lambda_0(t)dt\right)\right]^{\exp(\boldsymbol{\beta}^T \mathbf{z}_i)} \end{aligned} \quad (2.7)$$

となる.

2.2. ガンマ過程を用いたモデル

i 番目の被験者に対する観測値 (u_i, ω_i) と $(N_i(t), Y_i(t))$ は 1 対 1 に対応するため, $(N_i(t), Y_i(t), \mathbf{z}_i)$ を i 番目の被験者に対するデータとみなすことができる (服部, 2009). n 人のデータを $D = \{N_i(t), Y_i(t), \mathbf{z}_i\}_{i=1}^n$ と表すとき, $\boldsymbol{\beta}$ と $\Lambda_0(t)$ の同時事後分布 $p(\boldsymbol{\beta}, \Lambda_0(t) | D)$ は

$$p(\boldsymbol{\beta}, \Lambda_0(t) | D) \propto p(D | \boldsymbol{\beta}, \Lambda_0(t)) p(\boldsymbol{\beta}) p(\Lambda_0(t)) \quad (2.8)$$

で与えられる. ここに $p(\boldsymbol{\beta})$ は $\boldsymbol{\beta}$ の事前分布, $p(\Lambda_0(t))$ は $\Lambda_0(t)$ に対する確率過程事前分布, $p(D | \boldsymbol{\beta}, \Lambda_0(t))$ はデータ D が与えられたもとの尤度である. 打ち切りが non-informative であるとき, 被験者の生存時間が互いに独立として尤度は

$$p(D | \boldsymbol{\beta}, \Lambda_0(t)) = \prod_{i=1}^n \left[\prod_{t>0} I_i(t)^{dN_i(t)} \right] e^{-I_i(t)dt} \quad (2.9)$$

となる. (2.9)より, 尤度は平均 $I_i(t)dt$ をもつポアソン分布の核 $(I_i(t)dt)^{N_i(t)} \exp(-I_i(t)dt)$ に比例していることがわかる. 本稿では, Kalbfleisch (1978) に従い, 累積基準ハザード関数 $\Lambda_0(t)$ に対する事前分布としてガンマ過程 (gamma process) をとりあげる. ガンマ過程はガンマ分布に従って変化する確率過程であり, ポアソン分布の共役事前分布がガンマ分布であることを考えると, (2.9)において $\Lambda_0(t)$ にガンマ過程をおくことは自然な発想といえる.

ガンマ過程は以下のように記述される (Ibrahim et al., 2001).

(i) $Z(0) = 0$

(ii) $Z(t)$ は排反な区間において独立増分をもつ. 即ち, 任意の $t_1 < t_2 < \dots < t_n$ に対し, $Z(t_2) - Z(t_1), Z(t_3) - Z(t_2), \dots,$

$Z(t_n) - Z(t_{n-1})$ は互いに独立である.

(iii) $\alpha(0) = 0$ であり $[0, \infty]$ で左連続な増大関数 $\alpha(t)$ と $c > 0$ が存在するとき, 任意の $t > s$ に対して

$$Z(t) - Z(s) \sim Ga(\alpha(t) - \alpha(s), c)$$

である. ここに $Ga(\alpha, \beta)$ は平均 α/β , 分散 α/β^2 となるガンマ分布を表す.

上記 3 つの条件を満たす確率過程 $\{Z(t) | t > 0\}$ をガンマ過程とよび, 以降, 確率過程 $\{Z(t) | t > 0\}$ がガンマ過程に従うことを, $Z(t) \sim GP(\alpha(t), c)$ と表記する. いま, $\Lambda_0^*(0) = 0$ であり $[0, \infty]$ で左連続な増大関数 $\Lambda_0^*(t)$ に対し, $c > 0$ として $\Lambda_0(t) \sim GP(c\Lambda_0^*(t), c)$ であるとすれば, 微小時間 dt での累積基準ハザード関数の増分

$$d\Lambda_0(t) = \Lambda_0(t+dt) - \Lambda_0(t)$$

については, $d\Lambda_0^*(t) = \Lambda_0^*(t+dt) - \Lambda_0^*(t)$ として

$$d\Lambda_0(t) \sim Ga(c d\Lambda_0^*(t), c) \tag{2.10}$$

となる. (2.10) のとき $d\Lambda_0(t)$ の平均は $d\Lambda_0^*(t)$, 分散は $d\Lambda_0^*(t)/c$ となるため, c は $d\Lambda_0^*(t)$ への確信度を表す指標であることがわかる. $d\Lambda_0(t)$ を用いれば, 強度関数の増分 $I_i(t)dt$ は(2.6)より

$$I_i(t)dt = Y_i(t) \exp(\beta^T \mathbf{z}_i) d\Lambda_0(t) \tag{2.11}$$

となり, 尤度(2.8)は $dN_i(t), Y_i(t), d\Lambda_0(t)$ により表すことができる.

3. 文献での適用例

適用例として, Gehan (1965) の急性白血病データ (以下 Gehan のデータとよぶ) をとりあげる. 急性白血病的覚解維持のために 6-mercaptopurine (6-MP) とプラセボ群とを比較するために行われたランダム化比較試験であり, 6-MP 群 21 例, プラセボ群 21 例に対するデータが表 3.1 である. 白血病の再発をイベントとしており, 覚解状態から再発または打ち切りまでの時間が週を単位として測定されている.

表 3.1 : 急性白血病患者の再発までの時間 (週)

プラセボ群	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23
6-MP 群	6, 6, 6†, 6, 7, 9†, 10, 10†, 11†, 13, 16, 17†, 19†, 20†, 22, 23, 25†, 32†, 32, 34†, 35†

†: 打ち切り

この Gehan のデータに対し, 2.1.及び 2.2.で示した比例ハザードモデルをあてはめ, ベイズ流アプローチによる推論を行うための SAS プログラム例を以下に示す. このプログラム例は, DATA ステップによるデータハンドリングと MCMC プロシジャによるパラメータ推定の 2 つのパートで構成されている.

プログラム 3.1 は, DATA ステップによるデータハンドリングの一例である. 大橋・浜田 (1995) に倣い, 再発または打ち切りを表す変数 WEEK, 群を表す変数 GROUP (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 REMISS (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 を作成した. (2.9) に従い, MCMC プロシジャにて尤度を記載する必要があるため, SAS/STAT 9.3 (R) User's Guide, The MCMC Procedure の Example 54.14 を参考に, MCMC プロシジャで使用するためのデータセット A_GEHAN を用意した. Gehan のデータでは, イベント発現時間は重複を除くと 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15, 16, 17, 22, 23, 32 の 18 時点ある. そこでこれらイベント発現時間をもたせたデータセット PARTITION を用意し, 2 つのデータセット GEHAN と PARTITION から, 42 名の被験者に対し, 各イベント発現時点でのリスクセットを変数 m_Y に, 計数過程を変数 m_dN に, それぞれ入力したデータセット A_GEHAN を用

意した.

プログラム 3.1 : Gehan のデータハンドリング

```
data GEHAN ;
do drug = 0,1 ;
  do id = 1 to 21 ; input week remiss @@ ; output ; end ; end ;
cards ;
  1 1 1 1 2 1 2 1 3 1 4 1 4 1 5 1 5 1 8 1 8 1
  8 1 8 1 11 1 11 1 12 1 12 1 15 1 17 1 22 1 23 1
  6 1 6 1 6 1 6 0 7 1 9 0 10 0 10 1 11 0 13 1 16 1
  17 0 19 0 20 0 22 1 23 1 25 0 32 0 32 0 34 0 35 0
;
run ;
data PARTTITION ;
  input int_1-int_18 ;
  cards ;
  1 2 3 4 5 6 7 8 10 11 12 13 15 16 17 22 23 35
;
run ;
data A_GEHAN ; set GEHAN ;
if _n_ eq 1 then set PARTTITION ;
  array int[18] int_1-int_18 ;
  array Y[18]; array dN[18];
  do k = 1 to (dim(int)-1) ;
    Y[k] = (week - int[k] + 0.001 >= 0) ;
    dN[k] = Y[k] * ( int[k+1] - week - 0.001 >= 0) * remiss ;
  end ;
  output ;
drop k ;
run ;
data A_GEHAN ; set A_GEHAN ;
  array Y[*] Y; ;
  array dN[*] dN; ;
  do i = 1 to (dim(y)) ;
    y_val = y[i] ;
    dn_val = dn[i] ;
    int_index = i;
  output ;
```

```

end ;
keep Y_ : dN_ : int_ : drug int_index id ;
run ;
data A_GEHAN ; set A_GEHAN ; rename y_val = m_Y dn_val = m_dN ; run ;

```

プログラム 3.2 は、データセット A_GEHAN を用いて、MCMC プロシジャにより事後分布のモンテカルロ標本を生成するための SAS プログラム例である。WinBUGS 1.4 Examples Volume I の”Leuk: survival analysis Using Cox regression”を参考にした。Burn in を 1000、Burn in で除外された後のモンテカルロ標本におけるサンプリング間隔を 1、最終的にサンプリングされるモンテカルロ標本のサンプル数を 10000 個とした。回帰パラメータの事前分布には平均 0、分散 10^6 の正規分布を指定した。累積基準ハザード関数の増分 $d\Lambda_0(t)$ の事前分布にはガンマ分布(2.10)を指定した。Hobbs and Carlin (2008) に従い、 r を単位時間あたりのハザード、 dt を次イベント発現までの間隔として $d\Lambda_0^*(t) = r dt$ とおき、 $r = 0.1, c = 0.001$ と設定した。(2.9)のとおり、尤度がポアソン分布の核に比例することから、MODEL ステートメントにて平均 $I_i(t)dt$ をもつポアソン分布を指定した。煩雑さを避けるため、プログラム 3.2 では $d\Lambda_0(t)$ に対する初期パラメータの設定と事前分布の指定を SAS マクロ %dLprior により記述している。

プログラム 3.2 : 比例ハザードモデルを用いたベイズ流解析

```

%macro dLprior ;
array int[18] int_1-int_18 ;
%do i=1 %to 17 ;
parms m_dL0_&i. 1 ;
dL0star[&i.] = int[&i.+1]-int[&i.] ;
prior m_dL0_&i.~gamma(r* dL0star[&i.] * c, iscale=c) ;
%end ;
%mend dLprior ;
proc mcmc data = A_GEHAN
seed = 160721 nbi = 1000 nmc = 10000 nthin = 1 outpost = NLOUT ;
c = 0.001 ;
r = 0.1 ;
array dL0star[17] ;
array m_dL0[17] m_dL0_1-m_dL0_17 ;
parms beta 0 ;
prior beta ~normal(0.0,prec=1e-6) ;
%dLprior ;
m_ldt = m_y*exp(beta*drug)*m_dL0[int_index] ;
model m_dn~poisson(m_ldt) ;
run ;

```

表 3.2 に、回帰パラメータ β に対する推定結果として、事後平均、事後標準偏差、事後中央値、95%信用区間（事後分布の 2.5%点と 97.5%点）を示した。表 2.2 には示していないが、95%最高事後密度信用区間 (highest posterior density credible interval) は[-2.2894, -0.6258]であった。データセット NLOUT に出力されたモンテカルロ標本を用いることで、あるいはプログラム 3.2 にて MCMC プロシジャのオプションとして monitor=(HR) と追加し、MCMC プロシジャの中で beginnodata; HR=exp(beta); endnodata; と指定することで、対応するハザード比（プラセボ群に対する 6-MP 群のハザード比）の事後平均 0.2344、95%信用区間[0.089, 0.476]を得る。

表 3.2 : 比例ハザードモデルにおける回帰パラメータの推定結果

	Standard		Percentiles		
	Mean	deviation	Median	2.5	97.5
β	-1.5364	0.4187	-1.5213	-2.4223	-0.7427

比例ハザードモデルにおいて、生存関数は、累積基準ハザード関数を $\Lambda_0(t)$ として(2.7)より

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

となる。累積基準ハザード関数 $\Lambda_0(t)$ に $GP(c\Lambda_0^*(t), c)$ をあてはめたときには、ガンマ過程の定義より、時間軸 $(0, \infty)$ を、 $0 = a_0 < a_1 < a_2 < \dots < a_{J-1} < a_J = \infty$ と、 J 個の区間 $(0, a_1], (a_1, a_2], \dots, (a_{J-1}, a_J]$ に分割したとき、区間 $(a_{j-1}, a_j]$ での累積基準ハザード関数の増分

$$\Lambda_0(j) = \Lambda_0(a_j) - \Lambda_0(a_{j-1}) \quad (j=1, 2, \dots, J)$$

は平均 $d\Lambda_0^*(t)$ 、分散 $d\Lambda_0^*(t)/c$ をもつガンマ分布に従う。プログラム 3.3 は、上記に従い、データセット NLOUT に出力された 10000 個のモンテカルロ標本を用いて、各群の生存関数を推定する SAS プログラム例である。プログラム 3.3 により推定された生存関数の事後平均を図 3.1 に示した。

プログラム 3.3 : GP を用いた比例ハザードモデルによる生存関数の推定

```
proc sql ;
  create table SURVE as
  select * from NLOUT, PARTTITION ;
quit ;
data SURVE ;set SURVE ;
array int[18] int_1-int_18 ;
array aa[17] m_dL0_1-m_dL0_17 ;
do int_index = 1 to 17 ;
  t = int[int_index] ;
  m_dL0 = aa[int_index] ;
  output ;
end ;
keep lteration int_index t beta m_dL0 ;
run ;
data SURVE ;set SURVE ;by lteration ;
```

```

keep iteration t S_placebo S_6MP ;
if first.iteration = 1 then m_H0 = 0 ;
  m_H0+m_dL0 ;
  S_6MP = exp(-m_H0)**exp(beta*1) ;
  S_placebo = exp(-m_H0)**exp(beta*0) ;
run ;

```

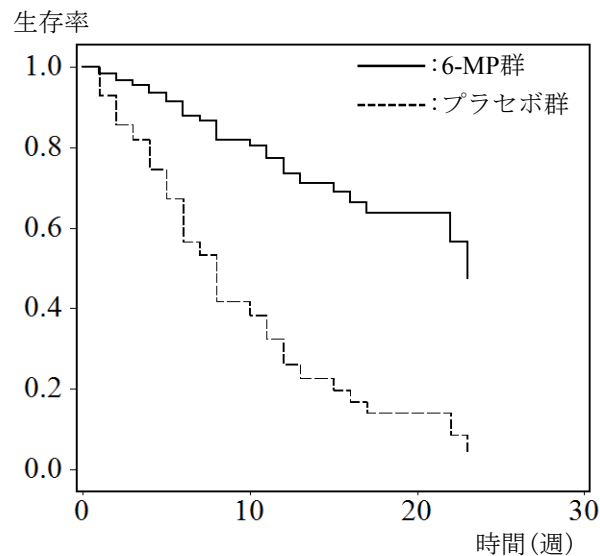


図 3.1 : Gehan のデータに対する GP-PH モデルによる生存関数

なお、基準ハザード関数に特定の事前分布をおくことなく、部分尤度に基づき推定された回帰パラメータを用いて生存関数を推定する場合には、データセット GEHAN を入力データセットとし、PHREG プロシジャにて BAYES ステートメントにより回帰パラメータの事前分布を指定し、BASELINE ステートメントにより各群に対する生存時間を出力させることができる。ただし、デフォルトの設定では共変量の平均値に対する生存時間のみが出力されるため、BASELINE ステートメントのオプション COVARIATE= で予め用意した各群を表す共変量の値をもたせたデータセットを指定する必要がある。プログラム 3.4 にその一例を示した。BAYES ステートメントのオプション COEFFPRIOR オプションを用い、回帰パラメータの事前分布に平均 0、分散 10^6 の正規分布を指定している。プログラム 3.2 と同様、Burn in を 1000、Burn in で除外された後のモンテカルロ標本におけるサンプリング間隔を 1、最終的にサンプリングされるモンテカルロ標本のサンプル数を 10000 個とした。

プログラム 3.4 : 部分尤度に基づく比例ハザードモデルによる生存関数の推定

```

data Prior ;
  input _TYPE_ $ drug ;
cards ;
Mean 0
var 1e+6

```



```

;
run ;
data Cov_vals ; do drug = 0,1 ; output ; end ; run ;
proc phreg data = GEHAN ;
  model week*remiss(0) = drug ;
  bayes seed = 160721 coeffprior = normal(input = Prior)
    nbi = 1000 nmc = 10000 thin = 1 outpost = Post ;
  baseline out = Estsurv survival = Surv
  covariates = Cov_vals /nomean ;
run ;

```

4. おわりに

ベイズ流解析では、パラメトリックな推論がよく用いられている。パラメトリックな推論では、あてはめる統計モデルにおける分布関数を有限次元のパラメータにより表現する。有限個のパラメータに対して事前分布を設定し、ベイズの定理を用いて得られる事後分布からパラメータの推論を行う。分布関数が有限個のパラメータで定式化できないとき、ノンパラメトリックな推論となる。ノンパラメトリックな推論では、通常、確率過程を用いて分布関数を表現する。得られたデータを確率過程からのサンプルパスと捉え、あてはめた確率過程からの経路によって決定される確率分布に従うと考える。Gehan のデータに対する適用例では、時間軸をイベントが発現した時点で分割し、各区間での累積基準ハザード関数の増分がガンマ分布に従うものとしてパラメータを推定した。このように、ノンパラメトリックとは、パラメータをおかないことではなく、無限個のパラメータをもつという意味である。想定したモデルでは無限個のパラメータであっても、得られたデータは“有限”であるため、得られたデータに対して解析するには有限個のパラメータでよく、有限個のパラメータにおとしこむことができれば、事後分布の計算が可能となる。

Kalbfleisch (1978) は、累積基準ハザード関数にガンマ過程やディリクレ過程をあてはめることの問題点として、確率 1 で累積基準ハザード関数が離散的な値をとること、隣接する区間で累積基準ハザード関数の増分を独立とみなすことをあげている。Dykstra and Laud (1981) は、基準ハザード関数の増分に拡張ガンマ過程をあてはめる方法を提案した。このことにより、上記に示した、ガンマ過程やディリクレ過程を累積基準ハザード関数にあてはめたときの問題点を回避できる。基準ハザード関数に対するモデリングとしては、本稿で取り上げたガンマ過程を用いる以外にも多数存在し、それらについて Ibrahim et al.(2001)によって紹介されている。こうした方法の多くは、1)基準ハザード関数に厳格なパラメトリック仮定をおかない、2)基準ハザード関数に対する推測がなされる、3)適当な事前分布を選ぶことにより推定された基準ハザード関数はもっともらしい滑らかな曲線となる、といった利点を有している (Dey and Rao,2005)。

本稿では、ベイズ流アプローチとして、累積基準ハザード関数に対しガンマ過程をあてはめた比例ハザードモデルをとりあげ、SAS を用いてこのモデルをあてはめ解析を行うための具体的な方法を紹介した。比例ハザードモデルに対するベイズ流ノンパラメトリックな方法について検討する際に、本稿がその一助になれば幸いである。

参考文献

- [1] Cox,D.R.(1972).Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society, Series B.* **34**:187–220.
- [2] Dey,D.K. and Rao,C.R. (2005). *Bayesian Thinking, Modeling and Computation*. Elsevier, Amsterdam. 繁梶算男・岸野洋久・大森裕浩監訳(2011). ベイズ統計分析ハンドブック. 朝倉書店.
- [3] Dykstra,R.L. and Laud,P.W. (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics.* **9**, 356-367.
- [4] Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrary single-censored samples. *Biometrika.* **52**: 203–223.
- [5] Hobbs, B.P. and Carlin, B.P. (2008). Practical Bayesian design and analysis for drug and device clinical trials. *Journal Biopharmaceutical Statistics.* **18**:54-80.
- [6] Ibrahim, J.G., Chen,M.H., Sinha,D. (2001). *Bayesian Survival Analysis*. Springer-Verlag, New York.
- [7] Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society: Series B.* **40**:214–221.
- [8] SAS Institute Inc. (2008). *SAS/STAT(R) 9.2 User's Guide*, Cary, NC, USA: SAS Institute Inc.
- [9] SAS Institute Inc. (2011). *SAS/STAT(R) 9.3 User's Guide*, Cary, NC, USA: SAS Institute Inc.
- [10]Sinha, D. Ibrahim, J. G. and Chen, M. H. (2003). A Bayesian justification of Cox's partial likelihood. *Biometrika.* **90**(3): 629–641
- [11]WinBUGS – The BUGS Project. WinBUGS 1.4, Examplpes Volume I, Leuk: survival analysis Using Cox regression. Available at http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS_Vol1.pdf.
- [12]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.
- [13]黒田普吾 (2010). Cox 回帰における Bayes 推定と PHREG プロシジャ.日本 SAS ユーザー総会 論文集,23-44.
- [14]服部聡 (2009). 生存時間解析におけるセミパラメトリック推測とその周辺. 統計数理. **57**(1):119-138.
- [15]中村剛 (2001). Cox 比例ハザードモデル. 朝倉書店.

連絡先

E-mail: yada-s@a2healthcare.com