$c=1, 2, \ldots, S,\ i=1, 2, \ldots, M,\ k=1, 2, \ldots, T,$

$Y_{ci}(T_k)$: T_k 時点に於いて c 番目の調査会社の得た i 番目のカテゴリの占有率のデータ,

β_{ci} : c 番目の調査会社の i 番目のカテゴリに対する相対的な調査特性,

$V_{ci}(T_k)$: 不規則成分（分散共分散構造は省略),

$$X_i(T_k)=2X_i(T_{k-1})-X_i(T_{k-2})+W_i(T_k),$$
$$i=1, 2, \ldots, M,\ k=3, 4, \ldots, T.$$

本報告では，不規則成分の分散共分散構造として，単純標本抽出とした時の標本誤差に比例した構造を採用した．それは，単純標本抽出を基準として，揺らぎの大きさを測ろうとしたからである．本報告の分散共分散構造は一つの例に過ぎず，原則として，退化した分布の分散共分散構造であれば何でも構わないのは言うまでもない．本報告の本質的な部分は，揺らぎ乗数を導入し，揺らぎの大きさをフレキシブルに推定しようとする所にある．

以上のモデルのパラメータの推定及びトレンドの予測は，赤池のベイズ手順に沿って行なう．参考として，揺らぎの大きさの推定の為のモデルの適用例を，図にて掲載しておく．図1はデータで，上から，内閣を支持する，支持しない，わからないと答えた者の占有率の時系列を表わす．そして，図2が予測されたトレンドを表わす．この時，α の推定値は 3.49 であった．

参 考 文 献

Akaike, H. (1980). Likelihood and the Bayes procedure, *Trab. Estadist.*, **31**, 143-166.

Kashiwagi, N. and Kishino, H. (1983). Estimation of the transition probabilities from the movement of the shares obtained by random sampling data, *Res. Memo.*, No. 254, The Inst. Statist. Math., Tokyo.

Kashiwagi, N. and Kishino, H. (1984). Estimation of the trend and fluctuation of time-series data collected with sample surveys, *Res. Memo.*, No. 282, The Inst. Statist. Math., Tokyo.

生存時間分布関数の推定とベイズモデル

統計数理研究所　鎌　倉　稔　成

1. は じ め に

生存時間分析は，医学・薬学および信頼性工学の分野において，近年，急速に脚光を浴びつつある統計的手法の１つである．とりわけ，Cox (1972)の開発した比例ハザードモデルは，背景因子などの共変量と寿命分布の関係を評価する標準的な方法になりつつある．統計解析の大きな目的の１つは，解析の対象とする集団になんらかの意味で同質性を仮定して，集団全体を支配している確率法則である寿命分布に対する情報を得ることにある．寿命分布の推定は通常の分布関数の推定と同様にして行われるが，大きく異なるのは，打ち切りを含む標本を対象とする点にある．経験分布に打ち切りデータを含む形で一般化したものに Kaplan and Meier (1958)の Product-limit estimate がある．これは普通 Kaplan-Meier の推定量と呼ばれ，生存時間分析の基本となるものである．

今，寿命データが n 個観測され，n 個のペアー (x_i, δ_i) $(i=1, \ldots, n)$ で表されるものとすれ

ば，Kaplan-Meier の推定量は次式のようになる．ただし，δ_i は打ち切りデータの場合には 0，そうでない場合には 1 を取るインディケーター関数とする．

$$\hat{S}(x)=\prod_{j:t_j<x}\left(1-\frac{d_j}{n_j}\right) \tag{1}$$

ここに，d_j は，$x_j(i=1,\ldots,n)$ の $\delta_i=1$ のものを順序付け，同順位のものを 1 つにまとめて $t_1,\ldots,t_k$ としたときの j 番目の死亡数（あるいは故障数）を示し，n_j は t_{j0} 時点で生存している個体数を表す．また，$0<x<t_1$ に対しては，$t_0=0$，$n_0=n$，$d_0=0$ と定義する．

また，生存時間分布関数 $S(x)$ を用いて $-d\log S(x)/dx$ と表現されるハザード関数に比例性を仮定した比例ハザードモデルにおいても，生存時間分布関数の推定問題はいくつか文献に見受けられる．たとえば，Breslow (1974) では，死亡のデータによって与えられる区間 $(t_{j-1}, t_j]$，$j=1,\ldots,k+1$ で基準ハザード関数が一定であることを仮定したモデルによって生存時間分布関数の推定を行っている．ただし，$t_{k+1}=\infty$ としている．比例ハザードモデルを

$$\lambda(t)=\lambda_0(t)\exp(z\beta) \tag{2}$$

とし，$\lambda_0(t)$ に対して，

$$\lambda_0(t)=\lambda_j \quad t_{j-1}<t\leqq t_j \quad (j=1,\ldots,k+1) \tag{3}$$

とすれば，λ_j の推定値は，最尤法によって次のように求められる（フィックスした β に対して）．

$$\hat{\lambda}_j=\frac{d_j}{(t_j-t_{j-1})\sum_{i\in R(t_j)}\exp(z_i\,\beta)} \quad (j=1,\ldots,k) \tag{4}$$

β に関しては次式を最大にするようにして推定され，これを (4) に代入し，ハザード関数と生存時間分布関数との関係式より生存時間分布関数の推定量が計算される．

$$L(\beta)=\prod_{j=1}^{k}\frac{\exp(\sum_{i\in D_j}z_i\beta)}{\{\sum_{i\in R(t_j)}\exp(z_i\beta)\}^{tj}} \tag{5}$$

ここで，t_j 時点においてのみ質量を持つ分布を考えると，

$$\hat{S}^*(t_j)=\prod_{l=1}^{j-1}\left[1-\frac{d_l}{\sum_{i\in R(t_l)}\exp(z_i\hat{\beta})}\right] \tag{6}$$

という推定量が得られる．(6) 式が多くの生存時間分析のプログラムパッケッジ(SAS, BMDP, SURVREG) に用いられている推定量である．また，Prentice and Kalbfleish (1979) の与えた，繰返し計算を利用した推定法もある．両者は漸近的に等価な推定量であることが知られている．これまでに議論した推定量は，いずれも Kaplan-Meier の推定量をベースにした離散的な推定量であったが，ここでは滑らかな生存時間分布関数を推定するための統計的方法について議論を与えることにする．

2. *B* スプラインを利用した推定法

スプライン関数は，区分的な多項式であり，プロットされた点を通る滑らかな曲線を得るために用いられている．スプライン関数は歪エネルギー最小を近似的に実現するという最適性を備えており，種々の応用が考えられている（市田・吉本，1979）．生存時間分布関数の推定の問題にスプライン関数を持ち込んだのは，Klotz (1981) である．彼は，B スプラインを利用して

滑らかな推定量を得ることに成功している．次節でこの推定方法にベイズ的手法を組込むために，B スプラインを利用した生存時間分布関数の推定法について簡単にふれることにする．r 次の B スプラインを考え，節点を，

$$\tau_{-r+1},\ \tau_{-r+2},\ \ldots,\ \tau_0,\ \tau_1,\ \ldots,\ \tau_k,\ \tau_{k+1},\ \ldots,\ \tau_{k+r-1}$$

とし，基準化された B スプラインを

$$N_{jr}(t)=(\tau_{j+r}-\tau_j)g_r(\tau_j,\ \tau_{j+1},\ \ldots,\ \tau_{j+r}\ ;\ t) \tag{7}$$

で定義する．ただし，$g_r(\cdot)$ は次式のように与えられる．

$$\begin{aligned}
&g_r(\tau_j\ ;\ t)=[\tau_j-t]_+^{r-1}\\
&g_r(\tau_j,\ \tau_{j+1}\ ;\ t)=[g_r(\tau_{j+1}\ ;\ t)-g_r(\tau_j\ ;\ t)]/(\tau_{j+1}-\tau_j)\\
&\quad\vdots\\
&g_r(\tau_j,\ \tau_{j+1},\ \cdots,\ \tau_{j+r}\ ;\ t)=[g_r(\tau_{j+1},\ \cdots,\ \tau_{j+r}\ ;\ t)-g_r(\tau_j,\ \cdots,\ \tau_{j+r-1}\ ;\ t)]/\\
&\qquad\qquad(\tau_{j+r}-\tau_j)
\end{aligned}$$

Klotz (1981) は,

$$\tau_{-r+1}=\tau_{-r+2}=\cdots=\tau_0=0\leqq\tau_1\leqq\cdots\leqq\tau_k=\tau_{k+1}=\cdots=\tau_{k+r-1}$$

とし，ハザード関数に対して

$$h(x)=\sum_{j=-r+1}^{k-1}\theta_j N_{jr}(x) \tag{8}$$

のモデル化を行って，最尤法によって θ_j をイテラティブに解く方法を与えている．特に節点を死亡時点に合わせ，$r=2$ としたときの累積ハザード関数の推定量は,

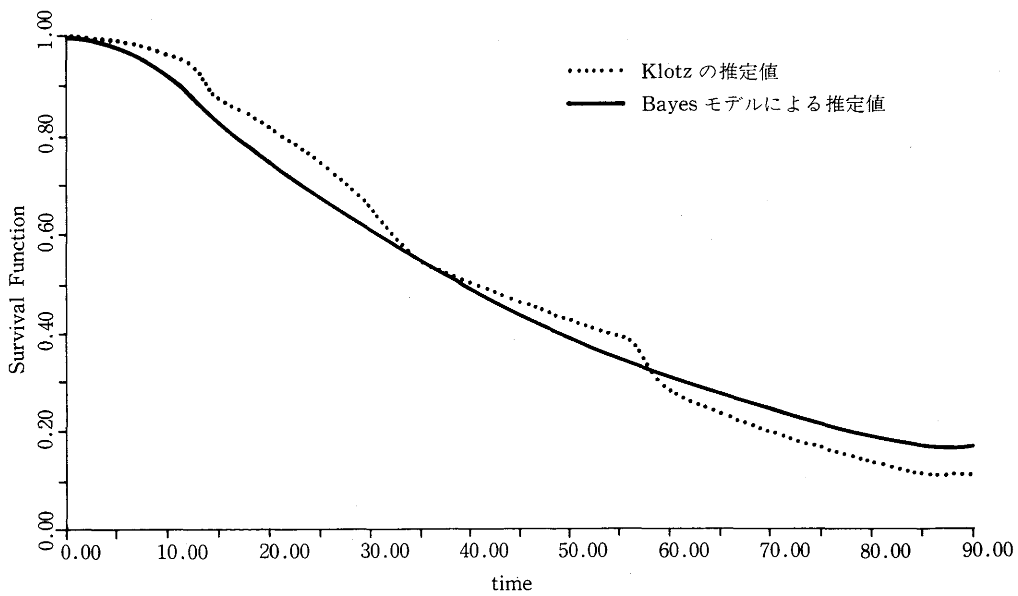

図 1.　Klotz の推定値と Bayes モデルによる生存時間分布関数の推定値（データは Klotz (1981) による）

$$\tilde{H}(x)=\sum_{l=1}^{k}\{m_l\sum_{j\geq l-1}N_{j3}(x)/\sum_{i=1}^{n}\sum_{r\geq l-1}N_{r3}(x_i)\} \tag{9}$$

で与えられる．$\exp\{-\tilde{H}(x)\}$ が生存時間分布関数の推定量となる．この推定量は一致性を持つことが証明されている．しかしながら，実際に推定された図を見ると，滑らかさが十分とは言えない．そこで次節では，滑らかな推定量を得るために最近よく用いられているベイズモデルによってこの問題を扱う．

3. ベイズモデル

モデルを一般化するために，滑らかに変化するパラメータの部分とそうでない，通常のパラメータの部分との両方を含むモデルについて議論する．滑らかに変化するパラメータは，たとえば，時系列データのように時間軸上に整列したパラメータ，あるいは，空間上の位置関係を保存したパラメータに見られる．比例ハザードモデルの場合は，基準ハザード関数を構成するパラメータがこれに対応する．今，通常のパラメータを，

$$\beta_1=(\beta_{11},\ \dots,\ \beta_{1p_1}),$$

滑らかさの情報をベイズモデルによって組込むパラメータを，

$$\beta_2=(\beta_{21},\ \dots,\ \beta_{2p_2})$$

で示す．前節で扱った B スプラインのモデルでは，θ_j に関して時間的変動を少なくするためのモデルを入れるので，β_1 の入っていないモデルとなる．比例ハザードモデルでは β_1 に対応するものは，共変量の回帰係数のパラメータである．β_2 の滑らかさに関する情報をここでは次の事前分布として表現する．

$$\pi(\beta)=(2\pi)^{-\frac{p_1+p_2}{2}}|D^{-1}\Sigma D'^{-1}|\exp\{-\beta'(D^{-1}\Sigma D'^{-1})^{-1}\beta/2\} \tag{10}$$

ただし，D, Σ は以下のように定義される．

$$D=\begin{bmatrix} I_{p_1} & 0 \\ 0 & D_2 \end{bmatrix},\qquad D_2=\begin{bmatrix} 1 & & & \\ -1 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{bmatrix},\qquad \Sigma=\left[\begin{array}{cc|cc} \sigma_1^2 & 0 & & \\ & \ddots & & 0 \\ 0 & \sigma_{p_1}^2 & & \\ \hline & & \sigma_a^2 & \\ & 0 & & \\ & & 0 & \sigma^2 I_{p_2-1} \end{array}\right]$$

このベイズモデルでは，事前分布を決定するパラメータ，超パラメータの設定が重要な問題となる．超パラメータの決定に際しては，Akaike (1980) の ABIC 最小化の基準が有効であることが多くの応用例に見受けられる．ここではこの基準によって超パラメータを選択することにする．ABIC を評価する際には，$f\cdot\Pi$ の β に関する積分が必要になるが確率密度関数 f が特別な形をしていない限りは，一般には容易ではない．そこで，

$$g(\beta)=\ln f(x\,|\,\beta)$$

とおいて，$g(\beta)$ を β_0 の回りでテーラー展開した2次近似を用いた．

前節の B スプラインのモデルに Bayes モデルを組入れるには，$p_1=0$, $p_2=k+r-1$ の β_1 の退化したモデルを考えればよい．ただし，$\beta_1, \dots, \beta_{p_2}$ を $\theta_{-r+1}, \dots, \theta_{k+r-1}$ に対応させる．また，死亡時間の起った各時点の間隔が一様でないので分散共分散行列 Σ の σ を t_j-t_{j-1} で重み付ける．このようにして得られた推定量を図1に示す．

参考文献

Akaike, H. (1980). Likelihood and the Bayes procedure, *Bayesian Statistics*, (eds. J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith), Univ. Press, Valencia, Spain, 143-166.

Breslow, N.E. (1974). Covariance analysis of censored survival data, *Biometrics,* **30**, 89-99.

市田浩三・吉本富士市 (1979). スプライン関数とその応用, 教育出版.

Kaplan, E.L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations, *J. Amer. Statist. Assoc.,* **53**, 457-481.

Klotz, J. (1982). Spline smooth estimate of survival, *Survival Analysis* (eds, J. Crowly and R.A. Johnson), Institute of Mathematical Statistics, Hayward, California, Vol. **2**, 14-25.

Prentice, R.L. and J.D. Kalbfleisch (1979). Hazard rate models with covariates, *Biometrics,* **35**, 25-39.

討　論

田辺：時間間隔が規則的でないのにこうしたベイズモデルを使ってもあまりうまくいかないのではないか.

鎌倉：Normalize した B スプラインを用いているので良いと思います.

田辺：どのような Normalize か. 高さなのか. Volume なのか.

鎌倉：この式で示される意味です ((7)式).

尾形：積分形の分布関数でみるよりインテンシィティーで見た方が意味があるのでないか.

鎌倉：そうした要求があるのであれば, それは簡単に計算できます. ここでは生存時間分布関数の滑らかな推定値ということで議論を展開しました.

非定常モデルとエントロピー最大化

統計数理研究所　北　川　源　四　郎

1. まえがき

エントロピー最大化原理によれば, モデルの良さはエントロピーによって評価する. しかし, 実際には, エントロピーは直接求めることができないので観測値から推定する必要がある. よく知られている様に, 対数尤度はこのエントロピーの推定を目ざしたものと考えることができる. したがって, たとえば, ある船のモデルがいくつか考えられるとき, その船の実際の運動の記録が得られると, 尤度を求めることによってモデルの良し悪しが比較できる. しかし, ここに一つの問題が生じる. データの採録中に, 海象や気象の変化や海流, 変針の影響などによって船体運動の様子が徐々にあるいは急激に変化することがある. この様な場合, 固定されたモデルの良さは時間とともに変化しており, 尤度はモデルの良さの平均的な値に対応しているにすぎない. もし, 各時刻におけるモデルの良さが何らかの形で評価できれば, それにもとづいて, 各時刻でモデル選択を行なうといったより細かい解析が行なえるようになるであろう.

第2節では, 時間とともに変化するエントロピーの様子をシミュレーションの結果を用いて例示する. 第3節では, 時間とともに変化するエントロピーを推定するために, 区分化尤度と平滑化尤度を提案する. 第4節では, これらを用いて非定常時系列の局所的モデル構成を行なっ