

確率的死亡モデルとその応用

統計数理研究所 鎌 倉 稔 成
" 柳 本 武 美

(1983年9月 受付)

1. 序

生存時間データの研究では、一般にある個体ないしはアイテムの死亡あるいは故障までの時間が主たる解析の対象である。文字どりの死亡、故障に限らず、ある事象が生起するまでの時間に関する問題もその対象となる。たとえば、喘息の発作のデータでは、患者の発作の時間間隔が計算され、この時間間隔を生存時間とみなす再生過程によるモデリング (Whittemore・Keller¹(1979), 柳本・鎌倉 (1982)) ができる。一方、入院加療した結核患者の再発の評価の問題では、治癒した時点から再発の時点までの時間を考えることにより生存時間の解析の手法を適用することができる。

事象の生起の問題をとらえるとき、事象は起きる起きないの2様であり、2値データとして考えられるのが通常である。これに対して機械や部品の故障では、故障という最終事態だけでなく故障に到るプロセスとしての劣化の度合も重要な問題である。また臨床医学の分野では患者の完治だけを問題にするのではなく、回復の度合をも考慮に入れた評価が重要である。従来の方法ではこうしたデータは適当なところで2値化して解析を行なう (たとえば、ある劣化のレベルを過ぎると故障とみなす) か、適当な時間軸で切り回復のレベルによって分割表を作ることが行なわれてきた。

痛みの軽減に有効な薬剤の評価のための臨床比較試験では、ある1時点における評価(著効, 有効, 無効)と治療との2元分割表による解析が通常行なわれている。しかしながら、データの収集の場ではより精緻なデータが集められるようになってきており、経時的な痛みの変化の記録も得られる。2元分割表による解析は、ある時点のみでの評価にすぎず、時間に関する情報は取り入れることができない。ところが、いま、A, B, 2種類の治療法によって図1のように時間を追って痛みの程度が軽減するものとすれば、 t_0 時点での痛みの程度はほぼ同じであっても、A治療法の方が効果があると見るのが自然である。むろん t_0 をいろいろ変えることによりA治療法の方が効果ありと判定することも可能かもしれないが、どこできるか (t_0 をどこに取るか) に問題が残る。

これらの問題はいずれも時間軸上にとられた単調に変化する曲線の識別の問題としてとらえることができる。部品の劣化の問題ではある部品の曲線が時間軸上で他の曲線より上側にあれば優れている部品であると解釈される。他方、患者の回復曲線でもその回復の度合を1から引いた曲線を考えると下側の曲線に対応する患者は回復過程が他に比べて良好であるとみることができる。

本稿では、患者の回復の過程あるいは部品の劣化の度合を考慮に入れたモデルを考え、その過程に影響を与える要因を抽出する方法を与える。

2. 確率的死亡モデル

2.1. 死亡・故障 (failure) の概念の一般化

“死亡” または “故障” ということばは文字どおりの個体の死亡, 機械の故障という意味だけでなく, 関心となっているある事象が生起するという意味に拡張できるということを序で述べたが, これはもう1つの意味でさらに拡張が可能である. 打ち切りのない場合の生存時間データは一般に次のような形式で整理ができる.

$$(2.1) \quad v_i = \begin{cases} 1 & (0 \leq t < t_i) \\ 0 & (t_i \leq t), i=1, \dots, n \end{cases}$$

通常は v_i は 0 と 1 の 2 値データであるが, これをより一般的に考えて v_i に 0 と 1 との間の連続量を許すように拡張することが考えられる.

腰痛の経時的な軽減現象では治療前の痛みを 1 とし時間を追って得られる痛みの強さから 0 と 1 との間の腰痛軽減曲線が得られる. むろん痛みが取れるまで継続的に観察して (2.1) の形でデータを整理することも可能であるが, 観察時間が長くなったり患者に脱落があったりしてそのようなデータを得るのは困難である. また, 痛みの軽減だけでなく時間とともに病態が悪化あるいは回復するような現象でも $v_i \in [0, 1]$ の解釈が可能である. 工学の分野ではたとえば発光ダイオードの発光効率の時間的低減でもこうした曲線が得られる.

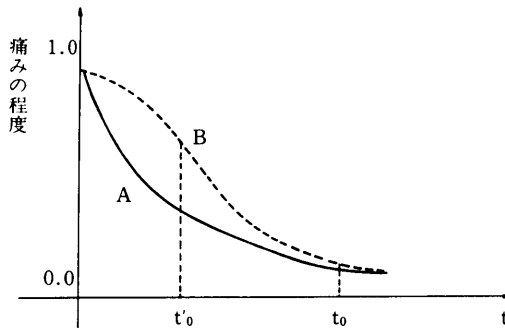


図1. 治療法の違いによる痛みの程度の変化

2.2. モデルとその評価関数

まずモデルを構成するために必要なデータ構造と条件についての記述を行なう. データはある定時点 t_j ($j=1, \dots, m$) において観測された患者の回復の度合または機械の劣化の程度であり, これを適当に基準化することによって

$$(2.2) \quad \begin{aligned} v_{i0} &= 1 & (i=1, \dots, n) \\ 0 \leq v_{ij} &\leq 1 & (i=1, \dots, n; j=1, \dots, m) \end{aligned}$$

と表わすことができる. 添字 i は観測される個体を識別するものである. ここでは v_{ij} に関しては次の仮定を置く.

$$(2.3) \quad v_{i0} \geq v_{i1} \geq \dots \geq v_{im} \quad (i=1, \dots, n)$$

この仮定は疾病の回復については逆もどりがなく, 一度劣化した部品では自然の状態では性能が良くなるようなことがないことに対応する. このように n 本の単調非増加な曲線の定時点における標本点が解析の対象となるデータである. この他, データとしては背景因子として考

えられる、個体の環境条件、個体特有の性質などの共変量と呼ばれる情報が与えられる。解析の目的は個体による曲線の不均一性を共変量で説明することであるが、その扱いは後の節に譲ることにして、まずは、総ての個体から得られる曲線が一様であることを仮定したときの全個体を用いた集団としての曲線を推定する方法を考える。

最初に曲線がステップ関数で表現される場合を想定する。ただし、関数の変化は観測時点 t_j ($j=1, \dots, m$) のみで起こるものとする。いまこのステップ関数で表わされる観測値がある個体の劣化の度合を示すものとして、劣化を次のモデルで考える。各個体はいくつかの類似の要素から成り立ち、この要素の故障数の割合によって劣化の度合が表現されるものとみなす。たとえば、全体の要素数の半分が故障すれば劣化の度合は0.5となる。要素の故障が自然に直ることがないとするれば劣化の度合は時間とともに進展し、単調非減少な曲線を与える。ここでは単調に変化する曲線を統一的に扱うために劣化の度合を1から引いたものを改めて劣化の度合と定義して単調非増加な関数として扱う。こうした定義によって劣化の度合は要素の生存時間分布のある推定値として与えられる。ここで要素の故障するまでの時間（生存時間）を T_{ik} ($i=1, \dots, n; k=1, \dots, K$) とし、それぞれが独立に密度関数 $f(t; \theta)$ を持つ連続分布に従っているとす。 K は劣化の度合を表現するために用いた量で未知であるが、個体ごとには変わらないものとする。曲線の各時点におけるステップ幅を

$$(2.4) \quad x_{ij} = v_{ij-1} - v_{ij} \quad (i=1, \dots, n; j=1, \dots, m)$$

とすれば x_{ij} はある確率分布に従う確率変数の実現値である。曲線があらかじめ決められた t_j 時点のみで観測されて得られたものであり、実際には (t_{j-1}, t_j) での変化が x_{ij} である（ただし、 $t_0=0$ ）とすれば、各 i に対して、 x_{ij} ($j=1, \dots, m$) は、

$$(2.5) \quad x_{ij} = [\text{number of } \{T_{ik}; T_{ik} \in (t_{j-1}, t_j], k=1, \dots, K\}] / K$$

と解釈される。

ここで $(x_{ij}; j=1, \dots, m)$ の同時分布、

$$(2.6) \quad L_{ci} = \Pr\{X_{ij} = x_{ij}; j=1, \dots, m\}$$

を考えると、 L_{ci} は T_{ik} ($k=1, \dots, K$) の順序統計量 $T_{(1)}, T_{(2)}, \dots, T_{(K \Sigma x_{ij})}$ の同時密度から構成される。よって、 $k_{i0}=0, k_{i\ell} = K \sum_{j=1}^{\ell} x_{ij}$ ($\ell=1, \dots, m$) と置いて

$$(2.7) \quad L_{ci} = \int_D \frac{K!}{(K - k_{im})!} \prod_{k=1}^{k_{im}} f(t_{(k)}; \theta) S(t_m; \theta)^{K - k_{im}} dt_{(k)},$$

ただし、 D は積分領域 $\{(t_{(1)}, \dots, t_{(k_{im})}); t_i < t_{(k_{i\ell+1})} \leq \dots \leq t_{(k_{i\ell+1})} \leq t_{\ell+1}, \ell=0, \dots, m-1\}$,

$$S(t_m; \theta) = \Pr\{T \geq t_m\} = \int_{t_m}^{\infty} f(t; \theta) dt.$$

また、 T_{ik} の実現値 t_{ik} が区間 $(t_{j-1}, t_j]$ に入るとき $f(t_{ik}; \theta)$ を $f(t_j; \theta)$ で置き替えることによって近似を行なうと、

$$(2.8) \quad L_{ci} \cong \frac{K!}{(K - k_{im})!} \prod_{j=1}^m \{f(t_j; \theta)(t_j - t_{j-1})\}^{K x_{ij}} S(t_m; \theta)^{K - k_{im}}$$

となる。 x_{ij} は i に関して独立な標本としているので $(x_{ij}; i=1, \dots, n; j=1, \dots, m)$ の同時分布 L_c は近似的に

$$(2.9) \quad L_c \propto \prod_{i=1}^n \prod_{j=1}^m f(t_j; \theta)^{K x_{ij}} S(t_m; \theta)^{K - k_{im}}$$

と表現できる。ここで、上式の右辺の対数を取り K で除したものを $\log L$ とすれば、

$$(2.10) \quad L = \prod_{i=1}^n \prod_{j=1}^m f(t_j; \theta)^{x_{ij}} S(t_m; \theta)^{1 - \frac{x_{ij}}{K}}$$

となり、未知変数 K の入らない形が得られる。

L がどのような性質を持つかについて観測時点を連続に取る場合について考えてみる。単調非増加をする曲線を $S_g(t)$ とする。ただし $S_g(t)$ は生存時間分布関数 (Survival function) の性質を持つ、つまり、 $G(t)=1-S_g(t)$ が分布関数となっているものとする。簡単のため、 $n=1$ の場合を考え、添字 i は省いて議論する。 V_j は $S_g(t_j)$ であり、(2.10)式は

$$(2.11) \quad L_1 = \prod_{j=1}^m \{f(t_j; \theta)^{\Delta S_g(t_j)}\} S(t_m)^{S_g(t_m)}$$

と表現できる。ただし、 ΔS_g は

$$\Delta S_g(t_j) = S_g(t_{j-1}) - S_g(t_j)$$

である。ここで分割 $\Delta^* : 0 = t_0 < t_1 < \dots < t_m = a$ を考え、 $d(\Delta^*) = \text{Max}\{t_k - t_{k-1} : k=1, \dots, m\}$ として $d(\Delta^*) \rightarrow 0, a \rightarrow \infty$ とすれば、(2.11)は

$$(2.12) \quad \log L_1 \rightarrow \int_0^{\infty} \{\log f(t; \theta)\} dG(t)$$

となる。(2.11)を最大にすることは要素の生存時間の密度関数を F とすれば、近似的に Kullback-Leibler の情報量の意味で分布 F と分布 G との距離を最小にすることに他ならない。

以下では K を除いた形の L を評価関数とした推定方式を議論する。 L は L_c と同一の推定量を与えるのは明らかであるが、最尤推定量の漸近的性質に基づく精度の評価はできない。 L_c よりも L をこれからの議論に用いることの利点は、(2.1)のタイプの通常のデータから得られる尤度が L と一致し、ここで述べる手法が利用できる点にある。つまり、 v_{ij} に 0, 1 しか許さない場合の尤度は L において、 $\{x_{ij}=0 \text{ or } 1\} \sum_{j=1}^m x_{ij}=1$ という特別な場合として与えられるのである。

3. 共変量の評価

前節で述べたように解析の目的はいくつかの単調非増加な曲線の不均一性を背景因子で説明することにある。故障解析の場合に変量として考えられるのは、たとえば部品の使用環境としての温度、湿度、圧力などが考えられる。痛みの軽減の観察などの臨床試験では、性別、年齢、治療方法等がある。

生存時間分布が複数の共変量の影響を受け一様性を仮定できないような母集団分布を扱う方法として回帰型のモデルが考えられる。その1つは母集団分布に、指数分布、ワイブル分布などのパラメトリックモデルを仮定し、その尺度パラメータを $\theta = z\beta$ ($z = (z_1, \dots, z_p)$, $\beta' = (\beta_1, \dots, \beta_p)$) として共変量の影響をみることである (Prentice (1973), Pike (1966))。他に Cox (1972) が提案した比例ハザードモデルのもとでのセミ・パラメトリックな方式がある。比例ハザードモデルではハザード関数を $\lambda(t; \beta, z)$ として

$$(3.1) \quad \lambda(t; \beta, z) = \lambda_0(t) \exp(z\beta)$$

のように共変量の影響を積の形でハザードに取り入れている。ただし、ハザード関数 $\lambda(t; \beta, z)$ は要素の生存時間 T の密度関数を $f(t; \beta, z)$, 分布関数を $F(t; \beta, z)$ としたとき $\lambda(t; \beta, z) = f(t; \beta, z) / (1 - F(t; \beta, z))$ である。また $\lambda_0(t)$ は基準ハザード関数と呼ばれる任意の正の関数である。この方法の長所は $\lambda_0(t)$ に特定の関数形を仮定しなくとも要因(共変量)の影響を示す回帰係数、 β の推定が可能である点にある。また比例性の仮定が多くの場合に合理的であるということも分かっている (Lawless (1982), p. 276)。

ここでは $\lambda_0(t)$ に適当な関数形を規定することなしにノンパラメトリックな方法を取ることにする。比例ハザードモデルを仮定して (2.10)式に代入してこれを LP と置けば、

$$(3.2) \quad LP = \prod_{i=1}^n \left(\prod_{j=1}^m \left[\lambda_0(t_j) \exp(z_{ij}\beta) \exp \left\{ - \left(\int_0^{t_j} \lambda_0(u) du \right) e^{z_{ij}\beta} \right\} \right]^{p_{ij}} \right) \\ \times \exp \left\{ - (1 - \sum x_{ij}) \left(\int_0^{t_m} \lambda_0(u) du \right) e^{z_{im}\beta} \right\}$$

が得られる。ここで、 $t_{j-1} < t \leq t_j$ で基準ハザードが一定、つまり、

$$\lambda_0(t) = \lambda_j \quad (j=1, \dots, m+1)$$

であり、 $t_0=0, t_{m+1}=\infty$ であるとすれば LP は次のように表現できる。

$$(3.3) \quad LP = \prod_{j=1}^m \{ \lambda_j^{x_{ij}} \exp(\sum_i x_{ij} z_{ij} \beta) \} \\ \times \exp \left\{ - \lambda_j (t_j - t_{j-1}) \left(\sum_{R(t_j)} x_{ji} e^{z_{ij}\beta} + \sum_i v_{im} e^{z_{im}\beta} \right) \right\}$$

ただし、

$$R(t_j) = \{k \mid t_k \geq t_j\}$$

である。

次にモデルの中の母数を求めるために $\partial \log LP / \partial \lambda_j = 0$ を満たす λ_j を $\hat{\lambda}_j$ とすると、

$$\hat{\lambda}_j = \frac{\sum_i x_{ij}}{(t_j - t_{j-1}) \left(\sum_i \sum_{R(t_j)} x_{ij} e^{z_{ij}\beta} + \sum_i v_{im} e^{z_{im}\beta} \right)}$$

が得られ、これを (3.2) に代入して整理すれば、

$$(3.4) \quad LP \alpha \prod_{j=1}^m \frac{\exp(\sum_i x_{ij} z_{ij} \beta)}{\left\{ \sum_i \sum_{R(t_j)} x_{ij} \exp(z_{ij}\beta) + \sum_i v_{im} \exp(z_{im}\beta) \right\}^{x_{ij}}}$$

となる。我々は、 β の推定に関心があるので (3.4) の右辺の最大化によって β の推定を行なう。

ここに述べた手法は Breslow [1] の比例ハザードモデルに対するアプローチを利用してゐる。したがってデータに適当な変換を行なえば通常の生存時間データの解析に開発されたコンピュータプログラムがそのまま適用できる。まず、適当な整数 U を探索して Ux_{ij} ($i=1, \dots, n; j=1, \dots, m$), Uv_{im} ($i=1, \dots, n$) が整数となるようにし、 LP に相当するものとして

$$(3.5) \quad \prod_{j=1}^m \frac{\exp(U \sum_i x_{ij} z_{ij} \beta)}{\left[U \left(\sum_i \sum_{R(t_j)} x_{ij} \exp(z_{ij}\beta) + \sum_i v_{im} \exp(z_{im}\beta) \right) \right]^{U \sum_i x_{ij}}}$$

を計算して最大になるように β を決定するのである。

4. 解 析 例

ここでは上述した確率的死亡モデルを利用して実際のデータ解析を行なう。

4.1. 脳卒中の回復過程

図2は代々木病院の二木氏の研究(二木(1983))で得られたある脳卒中患者3名の回復過程を示すデータである。脳卒中の障害は多岐にわたるが、機能障害の中核をなす片麻痺について、入院直後のBrunnstrom StageがIまたはIIの患者94人の6ヶ月間の回復過程がデータとして得られている。ここでは年齢と上肢麻痺、下肢麻痺の病型との回復過程に及ぼす影響をみるのが目的である。したがって共変量として年齢、型(下肢:0, 上肢:1)を導入する。Brunnstrom StageはIからVIまで(このデータではI, IIの区別は不能でまとめている)であり、観測時点は、0, 0.5, 1, 2, 3, 4, 6(単位:月)である。Brunnstrom Stageの評価ではそのStageが高いところにあるほど回復の過程は良好である。まずこのデータを確率的死亡モデルで扱えるよ

うに $[2, 6]$ を $[0, 1]$ に写し, 単調非増加となるように適当な変換を行なう. ここでは j 時点の Brunnstrom Stage の評価点を b_j としたとき $1 - (b_j - 2)/4$ の変換をする. 表 1 は確率的死

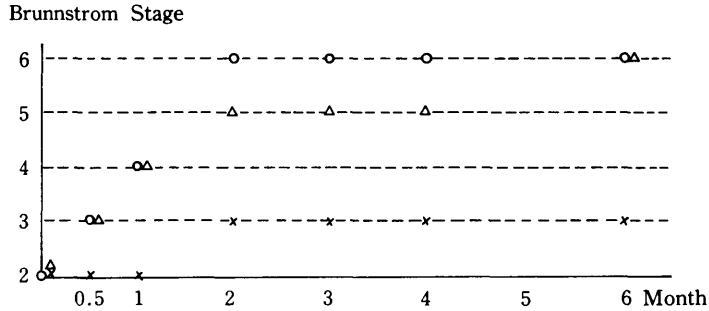


図 2. 脳卒中患者の回復過程

表 1. 脳卒中データの解析例

	Estimates	N.D.	P-values(1-sided)
1	-.03680	-2.73293	.00314
2	-.33635	-.86495	.19353

亡モデルによる解析結果である. 年齢の回帰係数の符号は負であり, 年齢が高いほど回復が遅いことが分かる. 型については上肢の方がなおりにくい, 年齢ほど回復過程には影響を及ぼしていない. これは二木氏の解析結果と同様な結論である.

4.2. 腰痛の軽減過程

図 3 は田無第一病院で得られたある腰痛患者の痛みの軽減過程である. データは治療前の腰痛を 10 としたときのアンケート調査による 10 段階評価で与えられる (吉田 (1983) 参照). 患者によっては単調性の仮定がくずれるような痛みの増加もみられるが, ここではこうしたデータは除去している. 表 2 は共変量を z_1 : 性別(1: 男 2: 女), z_2 : 年齢, z_3 : 型(1: 急性 2: 慢性), z_4 : 湿布剤(0: 未使用 1: 使用)としたときの回帰型の確率死亡モデルによって求められた回帰係数 β の推定値とその漸近正規性に基づく P 値である. 我々の評価関数が尤度でないため P 値の絶対的な大きさには意味がないが, P 値相互の相対的な大小は各係数の効き方の差の目やすとなる. この解析からは腰痛のタイプが痛みの軽減過程にかなり影響を及ぼしていることが分かる. 尚, このデータについては別に Kamakura et al. (1983) で報告されている.

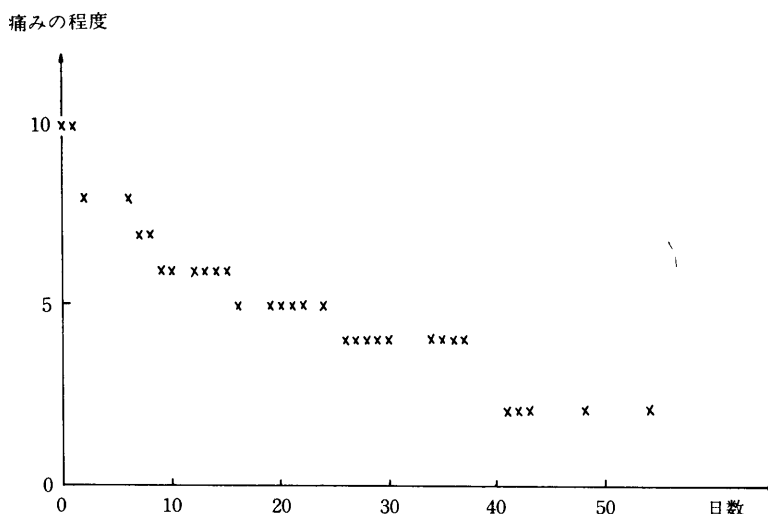


図3. 腰痛の寛解曲線の例

表2. 腰痛データの解析例

	Estimates	N.D.	P-values(1-sided)
1	-.07614	-.20662	.41815
2	.00822	.62351	.26647
3	-.76071	-1.96951	.02445
4	-.03359	-.05989	.47612

5. シミュレーション

前節では単調性の仮定がくずれている場合でも確率的死亡モデルによる解析が可能である例として腰痛患者の痛みの軽減曲線を扱ったが、ここではシミュレーションによりもう少し詳しくこの問題を議論する。

2群比較を考え、それぞれの曲線群の真の曲線は

$$(5.1) \quad \begin{aligned} S_1(t) &= \exp(-\lambda_1 t) \\ S_2(t) &= \exp(-\lambda_2 t) \end{aligned}$$

で表わされるものとする。また各群より n 個体ずつの標本を取り、 t_1, t_2, \dots, t_m 時点での m 回の観測を行なう。ここでは t_j 時点における曲線の観測値 v_{ij} が

$$(5.2) \quad \log \frac{v_{ij}}{1-v_{ij}} = \log \frac{v_{ij}^0}{1-v_{ij}^0} + \sigma e_{ij}$$

$$(5.3) \quad v_{ij}^0 = \begin{cases} \exp(-\lambda_1 t_j) & (1 \leq i \leq n) \\ \exp(-\lambda_2 t_j) & (n+1 \leq i \leq 2n) \end{cases}$$

で生成される場合を考える。 e_{ij} は $E(e_{ij})=0, E(e_{ij}^2)=1$ の互いに独立な正規分布に従うものとする。また共変量 z_{ij} には

$$(5.4) \quad z_{ij} = \begin{cases} 0 & (1 \leq i \leq n) \\ 1 & (n+1 \leq i \leq 2n) \end{cases}$$

を与える。実際のシミュレーションでは v_{ij} を 10 段階評価として小数第 2 位を四捨五入して小数第 1 位までとし、また $\lambda_1=1, \lambda_2=e^\beta$ ($\beta=0.5$) としてある。

表 3 は $\sigma=0$ の場合に観測時点数を変化させて β の推定値を計算したものである。ここでは観測時点 t を $[0, 2]$ の区間を m 等分して与えている。攪乱項がないので観測時点を十分多く取り、また v_{ij} の観測精度が高ければ β の推定値 $\hat{\beta}$ は 0.5 に近い値となる。これは Tsiatis (1981) によって証明された比例ハザードモデルにおける回帰係数の推定量の一致性からも予想されることである。

攪乱項のある場合には単調性の仮定を満足しないものが出て来るが、これは次のような手順によってデータの一部を削除する。 $Q_i^0 = \{v_{i1}, v_{i2}, \dots, v_{im}\}$ としたとき、 $v_{ij-1} < v_{ij}$, $j=2, \dots, m$ を満たす最小の j を持つ v_{ij} を取除き、 $Q_i^1 = Q_i^0 - \{v_{ij}\}$ とし、以下同様にして Q_i^2, Q_i^3, \dots を作り、 v_{ij} がすべて j に関して単調非増加となるまで繰返す。

表 4 は観測時点 t を $[0, 2]$ の 50 分割として攪乱項のある場合の β の推定値を示したものである。データの削除による偏りは免がれないが、 β は正として現われ、我々の検出目的である 2 群の上下関係は把握することができる。この場合は $\beta > 0$ であり、第 1 群 ($S_1(t)$ を基準にして得られた曲線群) の方が全体として上側にあることが見てとれる。

表 3. 観測時点の数と推定量の違い ($\sigma=0$)

m	5	10	20	50	100
$\hat{\beta}$	0.3699	0.4222	0.4394	0.5447	0.5447
P	0.0374	0.0211	0.0171	0.0055	0.0055

(試行回数100回の平均)

表 4. 2 群の比較 ($m=50$)

n	5	10	20	50	100	
$\sigma=1$	$\hat{\beta}$	0.3567	0.3551	0.3745	0.3806	0.3772
	P	0.2944	0.2230	0.1282	0.0337	0.0052
$\sigma=2$	$\hat{\beta}$	0.2302	0.2330	0.2470	0.2432	0.2479
	P	0.3577	0.3077	0.2320	0.1285	0.0493

(試行回数100回の平均)

6. 議 論

死亡あるいは故障の概念を一般化して確率的死亡(故障)モデルを提案し、単調非増加な曲線群の不均一性を種々の背景因子(共変量)で説明する方法を与えた。単調性の仮定をくずしているようなデータでも、単調性を壊している時点での観測データを取除くことによってこのモデルの当てはめが可能である。データの除去による情報損失は否めないが、シミュレーション結果からは共変量の不均一性を見る上ではあまり問題がないと言える。

他のモデリングの可能性としてはセミ・マルコフモデルが考えられる。Wu (1982) は Crowley・Hu (1977) で扱われている心臓移植のデータで 5 つの状態を定義して状態間の推移率(transition rate)に比例ハザードモデルを導入して解析を行なっている。その他 Lagakos et al. (1978) の研究があるが、いずれもよく定義されたあまり多くない状態を対象としている。状態数が多くなるとパラメータ数が多くなりそれだけ推定問題はむずかしくなる。これまで扱っ

てきた曲線群の識別の問題にセミ・マルコフモデルを用いるためには、 v_{ij} の取りうる値を各状態とみれば良い。しかしながら我々の関心事は各状態の推移ではなく曲線が全体として比較対象のものとは比べて上側にあるか下側にあるかの問題である。したがってセミ・マルコフのモデリングを行なうためには曲線全体の上下関係をどのようにパラメトライズするかを十分に議論しなければならない。今後の研究課題である。

この研究は、快くデータを提供して下さい、また討論をして頂いた田無第一病院、吉田貢氏、代々木病院、二木立氏によるところが大きい。感謝の意を表したい。また、査読者の皆様には大変有益な御意見を賜った。ここに厚く御礼を申し上げます。

参 考 文 献

- Breslow, R.E. (1974). Covariance analysis of censored survival data, *Biometrics*, **30**, 89-99.
- Cox, D.R. (1972). Regression models and life tables (with discussion), *J. R. Statist. Soc., B*, **34**, 187-220.
- Crowley, J. and Hu, M. (1977). The covariance analysis of heart transplant data, *J. R. Statist. Soc., B*, **72**, 27-36.
- Kamakura, T., Yanagimoto, T. and Yoshida, M. (1983). Construction of a stochastic failure model with application to data for pain relief, *Bulletin of the Biometric Society of Japan*, **4**, (to appear).
- Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *J. Amer. Statist. Ass.*, **53**, 457-481.
- Lagakos, S.W., Sommer, C.J. and Zelen, M. (1978). Semi-Markov models for partially censored data, *Biometrika*, **65**, 311-318.
- Lawless, J.F. (1982). *Statistical models and methods for lifetime data*, John Wiley & Sons, New York.
- 二木 立 (1983). 脳卒中患者の障害の構造の研究 ——片麻痺と起居移動動作能力の回復過程の研究——, *総合リハビリテーション*, **11**, 465-476.
- Pike, M.C. (1966). A method of analysis of certain class of experiments in carcinogenesis, *Biometrika*, **22**, 142-161.
- Prentice, R.L. (1973). Exponent survivals with censoring and explanatory variables, *Biometrika*, **60**, 279-288.
- Tsiatis, A. (1981). A large sample study of Cox's regression model, *Ann. Statist.*, **9**, 93-108.
- Whittemore, A.S. and Keller, J.B. (1979). Asthma and air pollution. A quantitative theory, *Proceeding of SIMS conference on energy and health*. N.E. Breslow and A.S. Whittemore (eds.), SIAM, Philadelphia.
- Wu, S. (1982). A semi-Markov model for survival data with covariates, *Mathematical Biosciences*, **60**, 197-206.
- 柳本武美・鎌倉稔成 (1982). 点過程モデルによる継続観察データの解析, *応用統計学*, **11**, 39-47.
- 吉田 貢 (1983). "腰痛症"とその周辺疾患に対する刺激点(ツボ)の選択 ——SSP刺激点の有効度の検討——, *SSP療法学術セミナー講演集*, 29-42.

A Stochastic Failure Model and Its Application

Toshinari Kamakura and Takemi Yanagimoto

(The Institute of Statistical Mathematics)

In the survival or the failure time analysis the failure time pertains to the time to the occurrence of some event. The failure is considered to be clearly judged whether it does occur or not. When we observe the process of pain relief or recovery from some disease, we can obtain a curve defined over time interval for each individual. Considering these curves as survival functions, a stochastic failure model is proposed, which permits us to utilize the proportional hazard model to explain the heterogeneity among curves by some covariates.

Key words: Failure time, Kullback-Leibler information, pain relief, proportional hazard model, survival function.