

## 脱落のある寿命データの数理解析

丸大 理 柳川 克

### 1. 序

1.1. 脱落のある寿命データ 生命科学における寿命データは動物実験データと臨床医学データに大別できる。両者の特長はつぎの通りである。

動物実験データ：例として発がん性が疑われている化学物質の発がん性テストを考えよう。実験動物はランダムに $k$ 群に分けられ、各群に対して相異なる量の化学物質が割りつけられ、ランダムに各個体に化学物質が投与される。厳密に管理の下で腫瘍発症までの日数が観察される。かかる状況の下では各群内の観測値は互いに独立、同一分布に従う確率変数の実現値とみなすことができる。したがって数学モデルを導入できる。しかしながら、実験途中で実験動物が他の原因で死亡したり、ある期間と経過したのも観察が打ち切られることがよく得られるデータは不完全データである場合が多い。

臨床医学データ：ある事象（治癒、がんの転移、死など）が発生するまでの期間が観察され、得られるデータに基づいて治療効果が比較される。このとき、患者が突然来院しなくなり、転院したりするにめ動物実験の場合と同様のデータの脱落、打ち切りの問題が頻繁に生じるほか、ヒトの生存には年齢、性別、生活環境などさまざまな因子が関与してくるからこれを考慮した数学モデルを立てデータ解析を行うことが必要がある。

これから2種類のデータは、寿命の長さを表わす変量のほか、 $p$ 個の変量  $Z = (Z_1, Z_2, \dots, Z_p)$  を導入することによって統一的に取り扱うことができる。例えば、動物実験データでは  $p=1$  とし、 $Z_1 = 1$  (群1),  $Z_1 = 2$  (群2),  $\dots$ ,  $Z_1 = K$  (群K) なるダミー変数  $Z_1$  を導入すればよい。また、臨床医学データでは、このほか、 $Z_2$  を性別、 $Z_3$  を年齢、 $\dots$  などを表わす変量に定めればよい。

2節が示すように、打ち切りは脱落の特別の場合として定式化できる。従って、本稿では脱落や打ち切りのあるデータを総称して脱落のあるデータとよぶことにする。脱落のあるデータは、少なくとも自体がある期間生存したという情報を持っている。これを捨てることは、せっかく得られた情報と利用しなればかりか、結果にバイアスを与えることにも

なりかわり。

1.2. 本稿の目的      Cox (1972) は脱落のある寿命データと、共変量を考慮しながら解析する能力を有法を提唱した。Cox法による生命科学データの解析は英米で大流行がある。Cox法の数学的側面を明らかにする研究も盛んになりつつある。特に従来のノンパラメトリック法の流れの中でCox法と位置付ける試みは面白そうであるが、部分的には成功している。本稿では、寿命分布の連続性がある場合に話を制限し、Cox法の数学的側面に焦点を当て、その特長を考察することを目的とする。

## 2. 定式化

脱落のある寿命データを数学的につぎのように定式化する。

・  $T_1^0, T_2^0, \dots, T_N^0$  ; 互いに独立な確率変数

$T_i^0$  は個体  $i$  の真の寿命に対応する確率変数である。脱落があるため観測可能とは限らない。

・  $F_i(t) = P\{T_i^0 \geq t\}$  ; 個体  $i$  の生存関数、連続性を仮定する。

・  $f_i(t) = -dF_i(t)/dt$

・  $\lambda_i(t) = \frac{f_i(t)}{F_i(t)} = \lim_{\Delta \rightarrow 0} \frac{P\{t \leq T_i^0 < t+\Delta | T_i^0 \geq t\}}{\Delta}$

$\lambda_i(t)$  はハザード関数とよばれる。個体  $i$  の時刻  $t$  における瞬間死亡率<sup>\*</sup>である。つぎの関係式が成立することは明らかである。

$$F_i(t) = \exp\left\{-\int_0^t \lambda_i(x) dx\right\} \quad (2.1)$$

- $U_1, U_2, \dots, U_N$  ;  $T_i^0$  に独立な確率変数,  $T_i^0$  と独立。
- $H_i(t) = P\{U_i \geq t\}$  ; 連続性を仮定する。
- $h_i(t) = -dH_i(t)/dt$

$U_i$  は個体  $i$  の脱落に関する確率変数である。

- $T_i = \min(T_i^0, U_i)$
- $\varepsilon_i = \begin{cases} 1 & ; T_i = T_i^0 \quad (\text{死亡}) \\ 0 & ; T_i > U_i \quad (\text{脱落}) \end{cases}$

脱落のある寿命データでは  $(T_i, \varepsilon_i)$ ,  $i=1, 2, \dots, N$  が観測可能である。問題は  $(T_i, \varepsilon_i)$ ,  $i=1, 2, \dots, N$  を観測して得られるデータに基づいて  $\lambda_i(t)$  の推定,  $\lambda_i(t)$  に関する仮説の検定, およびリスク要因の分析を行なうこと、言い換えれば、リスク要因を分析し、それによって調整し、それから生存関数  $F_i(t)$  を推定したり、検定したりすることである。ここで、 $F_i(t)$  として、工業製品の寿命試験の類似として、未知母数を持つ指数分布、ワイブル分布、対数正規分布など、 $H_i(t)$  として指

---

\* 本稿から事象の発生を“死亡”とよぶことにする。



$$L_c(t_1, \dots, t_N; \varepsilon_1, \dots, \varepsilon_N) = \prod_{j=1}^N [\lambda_j(t_j)]^{\varepsilon_j} \exp \left[ - \int_0^{t_j} \lambda_j(t) dt \right] \quad (2.2)$$

### 3. Cox の方法

Cox (1972) の方法は 2, の漸近的なアイデアよりなる。  
その一つは PH モデルの導入、他は部分尤度法の導入である。

#### 3.1. PH モデル

$$\ln(\lambda_i(t)/\lambda_0(t)) = \beta_1 z_{i1} + \dots + \beta_p z_{ip} = \beta' z_i \quad (3.1)$$

と置き、相対化された瞬間死亡率の対数と其変量  $z_1, \dots, z_p$  の線形結合で説明しようというものである。Cox (1958) のロジスティックモデルの応用に似てはならない。 $\lambda_0$  は  $Z$  の基準組に対する瞬間死亡率である。全人口平均瞬間死亡率などが利用できる場合、それと  $\lambda_0$  とすれば問題は簡単であるが、ここでは  $\lambda_0$  は未知である。 $z_i$  から平均を差引いた  $z_i - \bar{z}_i$  と置き、 $\bar{z}_i$  と置き、 $\bar{z}_i = 0$  のときの瞬間死亡率が  $\lambda_0$  であると理解すれば便利である。Cox は  $Z$  が時間に関係する変量を含むことを許しているが、モデルの一意性という点で若干の問題があるようである。ここでは、 $Z$  は時間に関係する項を含むものとしておく。

(3.1) を (2.2) 式に代入し対数部分尤度関数と求めると

$$\ln L_c(t, \varepsilon) = \sum_{j=1}^N [\varepsilon_j \beta' z_j + \varepsilon_j \ln \lambda_0(t_j)] - \sum_{j=1}^N e^{\beta' z_j} \int_0^t \lambda_0(x) dx \quad (3.2)$$

$\lambda_0$  に対応する密度関数を  $f_0$  とする。 $f_0$  として、 $\lambda_0$  に無関係な未知母数を含む指数分布、ワイブル分布、対数正規分布などを想定し、これらの未知母数と  $\beta$  に関して (3.2) 式を最大にすることによって  $\beta$  の推定量を求めることが出来る。実際松原・後藤 [1979] は  $f_0$  としてベキ正規分布を仮定し、適応的に分布型を推定しながら  $\beta$  の推定を行なう意欲的な方法を提案している。しかしながら、これらはつぎの2点でやっかいな問題を含んでいる。

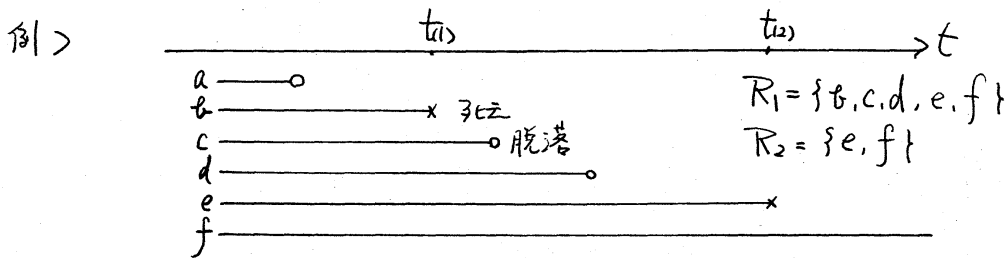
(1) 計算が (Cox 法にくらび) 極めて面倒である。

(2) PH モデルでは我々の関心は、才 1 に  $\beta$  に関する推定推定にある。このとき、 $\lambda_0$  は局外母数である。局外母数の存在するときの最尤推定量は、特にサンプル数が少ないとき、よく知られているように理論的に問題があり信頼性がない場合が多い。また、分布形の推定から生じる specification error の影響も無視ができない。

### 3.2 部分尤度法

まず、つぎのようにデータをいわゆる生命表形式にまとめることより始める。

- (3.3)  $\left\{ \begin{array}{l} t_{(1)} < t_{(2)} < \dots < t_{(n)} ; \text{ 死亡時刻} \\ i_j ; t_{(j)} \text{ で死亡した個体名} \\ R_j = R(t_{(j)}) ; \text{ 時刻 } t_{(j)} - 0 \text{ で生存しており、かつ観察} \\ \text{下にある個体名のリスト、リスク集合} \\ \text{という。} \end{array} \right.$



注1>  $R_i$  の連続性が仮定されているが、実際には死亡時刻は日で記録されているため同時刻に複数個の個体が死亡することもあり。このときの取り扱いも付録にゆかり、ここでは簡単のため重複がないものとしておく。

注2> 上のようデータをとめることは、区間  $(t_{(j)}, t_{(j+1)})$  で生じたすべての脱落を時刻  $t_{(j+1)} - 0$  で生じたと見做し、脱落に関する情報を覚えておくことに注意しよう。

(3.3) のデータ  $E$ 、 $R_j$  が所与で、集合  $R_j$  の中からランダムに1個個体を抽出すると  $i_j$  が抽出された個体が  $i_j$  であることと見做す。  $R_j$  の中から  $i_j$  が抽出される確率は

$$\lambda_{i_j}(t_{(j)}) / \sum_{i \in R_j} \lambda_i(t_{(j)})$$

これ  $j = 1, 2, \dots, l$  により掛け合わせるのか Cox の部分尤度関数である。即ち、部分尤度関数は次式で与えられる。

$$L_p(\beta) = \prod_{j=1}^l \frac{\lambda_{ij}(t_{ij})}{\sum_{i \in R_j} \lambda_i(t_{ij})}$$

PH モデル (3.1) を代入し、対数尤度を求めると

$$\ln L_p(\beta) = \sum_{j=1}^l \beta' z_{ij} - \sum_{j=1}^l \ln \sum_{i \in R_j} e^{\beta' z_i}$$

Cox の部分尤度法は、 $\beta$  に関する推定や検定を行う場合には  $\ln L_p(\beta)$  を普通の対数尤度関数と見做し、従来の最尤法と適用せよと主張する。上式は  $\lambda_0$  を含んでいない。よって、Cox の部分尤度法では、未知のハザード関数  $\lambda_0$  に依存することなく  $\beta$  に対する統計的推測を行うことができるわけである。

#### 4. 生存関数の推定

$\beta$  の推定値が求められれば、つぎに求められるのは共変量  $z$  で調整したときの生存関数の推定である。それには  $\lambda_0$  をデータから推定しておかねばならない。

(2.2) の条件ハザード  $L_c$  は、データが (3.3) の形に与えられるとき、注2で述べたことから注意して計算すると、つぎのように表現される。

$$L_c^* = \prod_{j=1}^q \lambda_{i_j}(t_{j-1}) \exp \left\{ - \sum_{j=1}^q \int_{t_{j-1}}^{t_j} \sum_{i \in R_j} \lambda_i(t) dt \right\} \quad (4.1)$$

PH モデル (3.1) を代し、さらに変形すると

$$L_c^* = L_P(\beta) \cdot L_D(\beta, \lambda_0)$$

と表わせる。ここに、 $L_P$  は Cox の部分尤度関数、 $L_D$  は次式で与えられる。

$$L_D(\beta, \lambda_0) = \left[ \prod_{j=1}^q \sum_{i \in R_j} e^{\beta' z_i} \lambda_0(t_{j-1}) \right] \exp \left[ - \sum_{j=1}^q \sum_{i \in R_j} e^{\beta' z_i} \int_{t_{j-1}}^{t_j} \lambda_0(t) dt \right]$$

$L_P$  には  $\lambda_0$  は含まれていない。よって、 $L_c^*$  と  $\lambda_0$  について最大にするには、 $L_D$  を最大にする  $\lambda_0$  を求めねばならない。本来なら  $L_c^*$  を最大にする  $(\beta, \lambda_0)$  が  $(\beta, \lambda_0)$  の (条件つき) 最大推定量かあるか、ここでのミソは、 $\beta$  の推定値  $\beta^*$  が部分尤度  $L_P(\beta)$  を最大にすることによって求められておくと考え、 $\beta = \beta^*$  と既知として  $L_D(\beta^*, \lambda_0)$  を  $\lambda_0$  について最大にすることによって  $\lambda_0$  の推定量を求めようと考えることである。これは Breslow (1972) のアイデアである。Cox (1972) も  $\lambda_0$  の推定法を提案しているが、理論的に根拠が薄弱であるように思われる。両者の比較はまだ行われていない。また、 $\beta$  が主役、 $\lambda_0$  が脇役の母数があるとして Breslow 法は直観的に極めて妥当な推定法と思われ、その正当性は如何にして保証できるであろうか。  $\lambda_0$  の推定に焦点を当てて

の問題を考察することは面白いそうである。これによって、Fraser (1968) が導いた marginal likelihood の考えが役に立ちそうである。実際、Kalbfleisch and Prentice (1973) は marginal likelihood の立場から Cox の部分尤度関数と理論的に導びく試みを行っている。

さて、

$$\lambda_0(t) = \lambda_i \quad t_{(i-1)} < t \leq t_{(i)}, \quad i=1, 2, \dots, l$$

とおく。このとき

$$L_D(\beta^*, \lambda_0) = \left[ \prod_{j=1}^l \sum_{i \in R_j} e^{\beta^* z_i} \lambda_j \right] \exp \left[ - \sum_{j=1}^l \sum_{i \in R_j} e^{\beta^* z_i} \lambda_j (t_j - t_{j-1}) \right]$$

$$\partial L_D(\beta^*, \lambda_0) / \partial \lambda_\alpha = 0 \quad \text{より}$$

$$\hat{\lambda}_\alpha = \left\{ [t_{(\alpha)} - t_{(\alpha-1)}] \sum_{i \in R_\alpha} e^{\beta^* z_i} \right\}^{-1}$$

を得る。これを (2.1) 式に代入すると生存関数の推定量

$$\hat{H}_0(t_{(k)}) = \exp \left\{ - \sum_{j=1}^k \frac{1}{\sum_{i \in R_j} e^{\beta^* z_i}} \right\} \doteq \prod_{j=1}^k [1 - \hat{\pi}_j]$$

$$\text{と得る。ここには、} \hat{\pi}_j = \left[ \sum_{i \in R_j} e^{\beta^* z_i} \right]^{-1}$$

特に  $\beta^* = 0$  のとき、 $\hat{H}_0$  は良く知られた Kaplan-Meier 推定量 (1958) と一致する。即ち、Breslow 法は共変量を調整した場合の Kaplan-Meier 推定量の一般化を提供する。

$F_i(t)$  の推定量は (3.1) 式より下記のようによえられる。

$$\hat{h}_i(t_{(k)}) = \exp\left\{-e^{\beta'x_i} \left[\frac{k}{\sum_{j=1}^k \frac{1}{\sum_{l \in R_j} e^{\beta'x_l}}}\right]\right\}$$

## 5. 部分尤度法の正当化

実は Cox (1972) では  $L_p$  は “条件，尤度” とよばれて  
いる。それかどのような意味で条件尤度であるのか、 $L_p$   
とよぶ，この尤度関数とみれば通常の最尤法と適用すればよい  
という根拠は何かには数多くの疑問が集中した。これに対  
して Cox (1975) は “条件，尤度” という言葉を取り消し、  
部分尤度という概念を導入して理論的正当化を行なった。そ  
の根幹はつぎの通りである。

$x_i = (R_i, t_i)$  とおく。事象は経時的に  $(x_1, i_1), (x_2, i_2),$   
 $\dots, (x_d, i_d)$  と観察されることより、(3.3) のデータを得る  
同時確率密度関数は

$$\begin{aligned} f_{\beta, \lambda_0}(i_1, \dots, i_d; R_1, \dots, R_d; t_1, \dots, t_d) &= f_{\beta, \lambda_0}(x_1) f_{\beta, \lambda_0}(i_1 | x_1) f_{\beta, \lambda_0}(x_2 | \\ i_1, x_1) f_{\beta, \lambda_0}(i_2 | i_1, x_1, x_2) &\dots = \prod_{d=1}^d f_{\beta, \lambda_0}(i_d | i_1, \dots, i_{d-1}; x_1, \dots, x_d) \\ \prod_{d=1}^d f_{\beta, \lambda_0}(x_d | i_1, \dots, i_{d-1}; x_1, \dots, x_{d-1}) \end{aligned}$$

こゝで

$$f_{\beta, \lambda_0}(i_d | i_1, \dots, i_{d-1}; x_1, \dots, x_d) = f_{\beta} (i_d | x_d)$$

$$f_{\beta, \lambda_0}(x_d | i_1, \dots, i_{d-1}; x_1, \dots, x_{d-1}) = f_{\beta, \lambda_0}(x_d | x_{d-1})$$

であることに注意する。明らかにつぎの対応が成り立つ。

$$L_c^* = f_{\beta, \lambda_0}(i_1, \dots, i_e; R_1, \dots, R_e; t_1, \dots, t_e)$$

$$L_P(\beta) = \prod_{\alpha=1}^g f_{\beta}(i_{\alpha} | x_{\alpha})$$

$$L_D(\beta, \lambda_0) = \prod_{\alpha=1}^g f_{\beta, \lambda_0}(x_{\alpha} | x_{\alpha-1})$$

簡単のため  $\beta \in R^1$  の場合について考える。

$$U_{\alpha} = \partial \ln f_{\beta}(i_{\alpha} | x_{\alpha}) / \partial \beta, \quad U = \sum_{\alpha=1}^g U_{\alpha} = \partial \ln L_P(\beta) / \partial \beta$$

$$J(\beta) = - \partial^2 \ln L_P(\beta) / \partial^2 \beta$$

$\beta^*$  :  $L_P(\beta)$  より求めらる最尤推定量

とおく。  $L_P$  を普通の密度関数と考え、通常の最尤法が適用できるためには、つぎの定理が、条件が与えられた分布  $f_{\beta, \lambda_0}(i_1, \dots, i_e; R_1, \dots, R_e; t_1, \dots, t_e)$  に対し成立すること示せばよい。

定理  $n \rightarrow \infty$  のとき

$$i) \quad U / \sqrt{J(\beta^*)} \rightarrow N(0, 1) \quad \text{in law}$$

$$ii) \quad (\beta^* - \beta) \sqrt{J(\beta^*)} \rightarrow N(0, 1) \quad \text{in law}$$

を示すには、  $f_{\beta, \lambda_0}(i_1, \dots, i_e; R_1, \dots, R_e; t_1, \dots, t_e)$  に関し、期待値  $E$  とする  $E(U) = 0$ ,  $V(U) = \sum_{\alpha=1}^g V(U_{\alpha})$  が成立すること、かつ

$$\textcircled{1} \quad E(U_{\alpha}) = 0$$

$$\textcircled{2} \quad E(U_{\alpha} U_{\alpha'}) = 0 \quad (\alpha \neq \alpha')$$

が成立すること示せばよい。  $\textcircled{1}$  については、  $U_{\alpha}$  は  $x_{\alpha}$  と所与としたときの条件密度によつて定義されるから

$E(U_\alpha | \lambda_\alpha) = 0$  は明らかである。よって、 $EU_\alpha = EE(U_\alpha | \lambda_\alpha) = 0$ 。

② について、 $\alpha < \alpha'$  のとき

$$E(U_\alpha U_{\alpha'}) = EE(U_\alpha U_{\alpha'} | U_\alpha = u_\alpha) = E[U_\alpha E(U_{\alpha'} | U_\alpha = u_\alpha)]$$

事象は経時的に生起するから  $\lambda_{\alpha'} \in \mathcal{P}$  とすると、 $\gamma$  は  $U_\alpha = u_\alpha \in \mathcal{P}$  とするとともに  $\mathcal{P}$  を含む。よって

$$E(U_{\alpha'} | U_\alpha = u_\alpha) = E(U_{\alpha'} | \lambda_{\alpha'}) = 0$$

これより ② は明らかである。

## 6. 部分尤度法の効率

$\beta$  に関する情報は明らかに  $L_D(\beta, \lambda_0)$  の方にも含まれている。にもかかわらずこれを考慮すると  $\beta$  の推定量が局外パラメータ  $\lambda_0$  に依存し、やっかいな問題が生じるため Cox はこれを切り捨てた。しかしながら切り捨てられる情報が無視できるほど大さければ部分尤度法の適用は躊躇せざるを得ない。

$p = 1$  のとき、 $L_C^*$ 、 $L_P$  より求められる Fisher の情報量は

$$I_C^*(\beta) = E\left[-\frac{\partial^2 \ln L_C^*}{\partial \beta^2}\right] = \sum_{j=1}^q E\left\{\frac{\sum_{i \in R_j} z_i^2 e^{\beta z_i}}{\sum_{i \in R_j} e^{\beta z_i}} \int_{t_{j-1}}^{t_j} \lambda_0(t) dt\right\}$$

$$I_P(\beta) = E\left[-\frac{\partial^2 \ln L_P}{\partial \beta^2}\right] = \sum_{j=1}^q E\left\{\frac{\frac{\sum_{i \in R_j} z_i^2 e^{\beta z_i}}{\sum_{i \in R_j} e^{\beta z_i}}}{\frac{\sum_{i \in R_j} z_i e^{\beta z_i}}{\sum_{i \in R_j} e^{\beta z_i}}} - \left[\frac{\sum_{i \in R_j} z_i e^{\beta z_i}}{\sum_{i \in R_j} e^{\beta z_i}}\right]^2\right\}$$

$L_C^*$  の替りに  $L_P$  を用いることより生じる情報の損失は

$$R(\beta) = \frac{I_p(\beta)}{I_c^*(\beta)}$$

で評価できる。  $R(\beta)$  を部分尤度法の効率とみなす。

特に 2 標本のとき

$$x_i = \begin{cases} 0 & : \text{標本 1} \\ 1 & : \text{標本 2} \end{cases} \quad \begin{aligned} n_0(t_j) &: R_j \text{ に属する標本 1 の回数} \\ n_1(t_j) &: R_j \text{ に属する標本 2 の回数} \end{aligned}$$

$$\lambda_0(t) \equiv 1, \text{ 即ち指数分布のとき}$$

とすると

$$R(\beta) = \frac{\sum_{j=1}^L E\{n_1(t_j)[t_j - t_{j-1}]\}}{\sum_{j=1}^L E\left\{\frac{n_0(t_j)n_1(t_j)}{[n_0(t_j) + n_1(t_j)e^\beta]^2}\right\}}$$

Kalbfleisch (1974) は脱離がないうちの場合、Efron [1977] は数学的な近似、Kay (1979) はモンテカルロ法によって  $R(\beta)$  の評価を行なっているが、彼らの方法は極めて複雑、難解で筆者の力不足のため要領よく紹介できない。理論的にもまだ多くの問題点が残っており、効率の評価は今後の課題である。彼らによって得られた結果はほぼおさるのようによまめることができる。

(i) 脱離がないうちとき、 $\beta = 0$  のときの評価

| 標本数 (N) | 10   | 20   | 40   | 60   | $\infty$ |
|---------|------|------|------|------|----------|
| $R(0)$  | 0.89 | 0.94 | 0.95 | 0.97 | 1        |

(ii)  $R(\beta)$  は  $\beta = 0$  のとき最大値をとり  $\beta$  が増加するにつれて減少する。

(iii)  $N \rightarrow \infty$  のとき、脱落のパターンが極端でなければ  $\beta = 0.5$  ぐらいのとき  $R(\beta) \geq 0.80$  と考えよう。多くの場合、この値は 95% 以上になる。  $N$  が有限のときは、上の表に  $N \rightarrow \infty$  のときの値をかけた値がほぼ  $R(\beta)$  の値に近いと見なしてよいのではないかと考えられている。

## 7. 今後の問題点

前節の2標本問題は

$$\text{矛1標本の分布: } F(t) = \exp\{-\int_0^t \lambda_0(x) dx\}$$

$$\text{矛2標本の分布: } G(t) = \exp\{-e^\beta \int_0^t \lambda_0(x) dx\}$$

であるとき、仮説  $H_0: \beta = 0$  を検定する問題にほかならない。 $\lambda_0$  は任意だから上のように  $F(t)$  に特別の型を指定しても意味はない。したがって、問題は矛1, 矛2標本の母集団分布を  $F, G$  とし、 $F, G$  が未知とすると、 $G = F^\psi$

( $\psi = e^\beta$ ) と表わされることから仮説:  $\psi = 1$  の検定問題に帰着する。この問題は、Lehmann<sup>(1953)</sup> の対立仮説に対する検定問題としてノンパラメトリック法では古くからよく研究されている。2標本問題に於いても同様の対応がとれる。従って Cox 法は脱落のあるデータに基づいて Lehmann 対立仮説を検定す

る方法とその特別の場合として提供する。他方、ノンパラメトリック法の枠内でも局所的最強の順位検定という基準に基づき脱着のあるデータに対する順位検定が研究されている(例えば Johnson and Mehrota (1972), Mehrota and Johnson (1976) など)。Cox の部分尤度法は極めて広範な問題に適用可能であるが、上のように問題を制限し、ノンパラメトリック法の立場から脚光を当て、その最適性を何らかの意味で数学的に証明することは可能であるように思われる。Cox 法を数学的に明確に位置づけることは今後に残された課題がある。

### 参考文献

- Cox, D.R. (1972) : Regression models and life tables (with discussion) J.R.S.S. B34, 187-220
- Breslow, N. (1972) : 上の論文の Discussion
- Fraser, D.A.S. (1968) : The structure of inference, New York, Wiley
- Kalbfleisch, J.D. and Prentice, R.L. (1973) : Marginal likelihood based on Cox's regression and life model. Biometrika 60, 267-278
- Kaplan, E.L. and Meier, P. (1958) : Nonparametric estimation from incomplete observation J.A.S.A. 53, 457-481
- Cox, D.R. (1975) : Partial likelihood, Biometrika, 62, 267-279

Kalbfleisch, J.D. (1974) : Some efficiency calculations for survival distributions.

Biometrika 61, 31-38

Efron, B. (1977) : The efficiency of Cox's likelihood function for censored data.

J.A.S.A. 359, 557-565

Kay, R. (1979) : Some further asymptotic efficiency calculations for survival data regression models. Biometrika 66, 91-96

Lehmann, E. L. (1953) : The power of rank tests. Ann. Math. Statist. 24, 23-43

Johnson, R.A. and Mehrota, K.G. (1972) : Locally most powerful rank tests for the two-sample problem with censored data. Ann. Math. Statist. 43, 823-831

Mehrota, K.G. and Johnson, R.A. (1976) : Asymptotic sufficiency and asymptotically most powerful tests for the two sample censored situation. The Ann. of Statist. 4, 589-596

松原義弘・後藤昌司 (1979) : センサー帰による生存時間データの解析, 応用統計シンポジウム予稿集「医学・生物学における統計的諸問題とデータ解析」

## 付録 [コンピュータ・プログラム化へのステップ]

### — 重複度のあつ一般の場合 —

#### (1) データの整理

| 死亡時刻     | 重複度      | 共変量       |           |     |           |
|----------|----------|-----------|-----------|-----|-----------|
| $t_{ij}$ | $m_{ij}$ | $z_{1ij}$ | $z_{2ij}$ | ... | $z_{rij}$ |

の output.  $z_{kij}$  より  $i$  の平均  $\bar{z}_k$  と差 (3) を,  $i$  を改

に  $z_{kij}$  と定めておく,  $k=1, 2, \dots, P$ .

$$(2) \quad \left. \begin{array}{l} t_{ij} \\ \text{重複, } t_{ij} \\ \vdots \\ t_{ij} \end{array} \right\} \begin{array}{l} \text{--- } z_{ij_1} \\ m_{ij} \text{ --- } z_{ij_2} \\ \vdots \\ \text{--- } z_{ij_{m_{ij}}} \end{array} \quad \Delta_{ij} = \sum_{l=1}^{m_{ij}} z_{ij_l} \quad \text{と求める.} \quad (A1)$$

リスク集合  $R_j$  の中から相異なる  $m_{ij}$  個の要素の組を抽出し、その各々についで、上の  $\Delta_{ij}$  を計算する。

$$R_j \left\{ \begin{array}{l} 0_{m_{ij}} \rightarrow \Delta_{ij_1} \\ 0_{m_{ij}} \rightarrow \Delta_{ij_2} \\ \vdots \\ 0_{m_{ij}} \rightarrow \Delta_{ij_{m_{ij}}} \end{array} \right\} \left\{ \begin{array}{l} R_j \text{ の要素の個数と } \gamma_j \text{ とおくと} \\ \text{この個数は} \\ (\gamma_j) \text{ 個} \end{array} \right. \quad (A2)$$

対数尤度関数は

$$\ln L_p(\beta) = \sum_{j=1}^J \beta' \Delta_{ij} - \sum_{j=1}^J \ln \left[ \sum_{k \in R_j}^* \exp \beta' \Delta_{jk} \right]$$

ここに  $\sum_{k \in R_j}^*$  は (A2) の  $(\gamma_j)$  個に属する和を意味する。

(3)  $\beta$  の推定、検定

$\beta = 0$  を初期値として、くり返し計算で  $\ln L_p(\beta)$  が最大にする  $\hat{\beta}$  を求める。  $\partial \ln L_p(\beta) / \partial \beta_k |_{\beta=\hat{\beta}}$ , Fisher の情報行列、分散共分散行列の推定値、標準化された  $\hat{\beta}$  などの output。  $\beta$  に関する仮説の検定は通常の方法に従う。或いは AIC 基準を用いる。

(4) 生存関数の推定

(3) より 変量の 個数  $p$  と  $\beta^* = \hat{\beta}$  を定め、それを用いる

データに用いる。

$$\hat{\pi}_j = \frac{m_j}{\sum_{k \in R_j} \exp(\beta^* z_k)}$$

$$\hat{H}_0(t_{(k)}) = \prod_{j=1}^K (1 - \hat{\pi}_j)$$

$\hat{H}_0$  の分散の 1 つの近似的な推定量は次式で与えられる。

$$\begin{aligned} V[\hat{H}_0(t_{(k)})] &\doteq \left[ 1 - \frac{m_k}{\sum_{l \in R_k} e^{\beta^* z_l}} \right]^2 V[\hat{H}_0(t_{(k-1)})] \\ &\quad + [\hat{H}_0(t_{(k)})]^2 \frac{m_k}{\left( \sum_{j \in R_k} e^{\beta^* z_j} \right)^2}, \quad k=1, 2, \dots \end{aligned} \quad (A.1)$$

よって、初期値  $V[\hat{H}_0(t_{(0)})] = 0$  より逐次に計算することにより  $\hat{H}_0$  の標準偏差を求め、生命表形式に  $t_{(k)}$ ,  $\hat{\pi}_j$ ,  $\hat{H}_0(t_{(k)})$ ,  $SE(\hat{H}_0(t_{(k)}))$ ,  $k=1, 2, \dots$  を output する。

11 ← 7 の層に層別されているとは各層ごとの操作を行い  $\hat{H}_0(t_{(k)})$  を求めればよい。

[補遺] (A.1) 式の導出方法)

筆者の知る限りにおいて  $\hat{H}_0$  の分散の推定量と与えられた文献は存在しない。(A.1) の導出方法を簡単に説明しておく。本稿の枠組で、重複度があるとき  $\ln L$  は次式で与えられる。

$$\ln L_D(\lambda_1, \dots, \lambda_K) = \sum_{j=1}^J \ln \sum_{i \in R_j}^* e^{\beta^* z_i} + \sum_{j=1}^J m_j \ln \lambda_j - \sum_{j=1}^J \sum_{i \in R_j} \lambda_j (t_i - t_{j-1}) e^{\beta^* z_i}$$

