

窓打切りされた再発事象のモデルとパラメータの推定方法

The models of recurrent events under the window censoring and their parameter estimation

経営システム工学専攻 阿部興

Industrial and Systems Engineering Ko ABE

1 序論

本研究はヒートシールの強度の分析から始まった。ヒートシールとは界面を熱で溶かして他の界面に接着する方式のシールを指す。ヒートシールの界面を顕微鏡写真によって観察すると、その界面は線分とみなすことができ、接着された界面と未接着の界面が交互に現れる。ある温度で接着したとき、界面がどのように接着され、どのように未接着であるか、その分布を知ることはヒートシールの品質を管理する上で重要である。しかし、界面は顕微鏡の視野の範囲でしか観測され得ない。その視野からは、界面の正確な長さをすべて知ることはできない。本論文では観測の可能な区間の範囲を「窓」と呼び、観測が限定された窓からなる場合の再発事象のモデルとそのパラメータの推定を論じる。

生存時間分析や信頼性工学などの分野では、上述のように観測範囲が制限されることから生まれる、データの不完全性にどのように対処するかが、かねてより一つの大きな問題とされてきた。Laslett (1982) は空間上の線分の長さの分布を推定する問題と、生存時間の分布を推定する問題の共通性を指摘している。人間や機械の寿命データを扱う際は、すべての対象が死亡や故障といったイベントの生起を迎えるまで観測を続けることは現実的でない。その場合、観測はある決まった区間内でのみ行われ、その区間の中でイベントが生起したか否かと、イベントの生起した時点を記録することになる。観測が終了するまでにイベントの生起を確認できなかった場合、右打ち切りが生じたという。右打ち切りに加えて、直前のイベントの生起が確認されず、イベントの生起までの時間がある時点からの余命として観測される場合、窓打切りされた (window-censored) という (Zhao, 2006, Zhu *et al.*, 2014, Rootzén & Zholud, 2016)。

Vardi (1982) は限定された区間内で観測された再生過程の再生間隔の分布について尤度を構成し、ノンパラメトリック最尤推定量を導いた。パラメトリックな推定に関しては、Zhao (2006) が詳しい。再生間隔の分布にワイブル分布、ガンマ分布、対数正規分布などを仮定し、窓打切りされた場合に対応した完全な尤度関数を与えている。

このように再生過程の研究はなされているが、ヒートシール界面の接着、未接着のように 2 種類のイベントが交互に起きる交代再生過程についての研究はほとんどない。交代再生過程は多くの分野で関心が持たれており、信頼性工学の分野では、稼働と休止の 2 状態を繰り返すシステムは交代再生過程を用いて記述される。Rootzén & Zholud (2016) では自動車の運転者が道路を見ている状態を 1、道路以外の箇所を見ている状態を 0 に対応させてモデル化し、状態 0 の長さの分布のパラメータに対して条件付き最尤法に基づく推定量を提案した。

2 対象とする確率過程

交代再生過程 (alternating renewal process) は 2 つの状態 (状態 0 と状態 1 とする) を交互に繰り返す再生過程の一般化である。

互いに相異なる正の連続型の確率分布関数, $F_1(x)$, $F_0(y)$ を考える. 大きさ n の確率変数列 X_1, \dots, X_n , Y_1, \dots, Y_n は互いに独立でそれぞれ $F_1(x)$, $F_0(y)$ に従う. $F_1(x)$, $F_0(y)$ はそれぞれ有限の平均 μ_1 , μ_0 を持つとする.

プロセスは状態 1 から始まり, 状態 1 を時間 X_1 で, 状態 0 を時間 Y_1 で過ごす. 続けて組の変数 $(X_2, Y_2), (X_3, Y_3), \dots, (X_n, Y_n)$ を繰り返し考える. 過程は状態 1 と 0 が交互に切り替わる. この過程を交代再生過程と呼ぶ. $Z_i = X_i + Y_i$ とする. 再生の起こった時間を $S_k = \sum_{i=1}^k Z_i$ と表し, 計数過程 $N(t) = \max\{k | S_k < t\}$ を構成する. プロセスでの状態を記録するために $V(t) = I\{S_{N(t)-1} + X_{N(t)} > t\}$ なる関数を導入する. ここで $I\{\cdot\}$ は指示関数であり, 不等式を満たすとき 1, さもなくば 0 の値を取る.

交代再生では状態 1 にいる確率の極限 ρ は,

$$\rho = \lim_{t \rightarrow \infty} \Pr\{V(t) = 1\} = \mu_1 / (\mu_0 + \mu_1). \quad (2.1)$$

また交代再生過程において,

$$S_{1,n} = \sum_{i=1}^{n-1} (X_i + Y_i) + X_n, \quad S_{0,n} = \sum_{i=1}^n (X_i + Y_i)$$

と置く. $Z_1(t)$ を以下のように, 過程が t 時点で状態 1 にあるときの次のイベントの生起までの待ち時間 (余命) を表す変数とする.

$$Z_1(t) = S_{1, N(t)+1} - t. \quad (2.2)$$

このとき, 以下が成り立つ (Beichelt & Paul, 2002).

$$G_1(x) = \lim_{t \rightarrow \infty} \Pr(Z_1(t) > x | V(t) = 1) = \frac{\int_x^\infty \{1 - F_1(u)\} du}{E[X_n]}. \quad (2.3)$$

過程が t 時点で状態 0 にあるときも同様, $Z_0(t)$ を過程が t 時点で状態 0 にあるときの余命を表す変数とすると, 以下を得る.

$$G_0(y) = \lim_{t \rightarrow \infty} \Pr(Z_0(t) > y | V(t) = 0) = \frac{\int_y^\infty \{1 - F_0(u)\} du}{E[Y_n]}. \quad (2.4)$$

$G_1(x)$, $G_0(y)$ は生存関数である. 対応する密度関数を $g_1(x)$, $g_0(y)$ と表す.

3 窓打切り状況下での再発事象のパラメータ推定

窓打切り状況下での再発事象のパラメータ推定には, 多くの場合最尤法に基づく方法が取られる. Laslett (1982), Vardi (1982), Zhao (2006) に共通する特徴は, 観測されるイベントの生起間隔を, 完全に観測され

た場合、窓の始点で打ち切られた場合、窓の終点で打ち切られた場合、窓の始点と終点の両側で打ち切られた場合の 4 通りに分類することで尤度関数を導出した点である。続く 4 章ではこれを踏まえ、窓打ち切りされた交代再生過程の尤度関数を導く。

4 窓打ち切りされた交代再生過程のパラメータの最尤推定

$V(t)$ を 0 または 1 の値を取る交代再生過程とする。我々は交代再生過程を区間 $[0, w]$ の窓を通じて観測する。交代再生過程の 1 つのサンプルパスに対して、窓はただ 1 つである。 τ_i ($i = 1, \dots, m$) を観測される状態 0 または 1 の長さとする。状態 0 の長さは正の連続型の分布関数 $F_0(\tau)$ を持ち、状態 1 の長さも同様、正の連続型の分布関数 $F_1(\tau)$ を持つとする。密度関数はそれぞれ、 $f_0(\tau)$, $f_1(\tau)$ で表す。また状態 0 の長さの分布と状態 1 の長さの分布は独立とする。 d_i を右打ち切りの有無を示すインジケータとし、 τ_i が右打ち切りされたとき 0、そうでなければ 1 の値を取るとする。 a_i を左打ち切りの有無を示すインジケータとし、 τ_i が観測開始の原点から始まるとき 0、そうでなければ 1 の値を取るとする。 s_i を状態を示すインジケータとする。また、 n を観測される窓の総数とし、 n_1 を観測が状態 1 からはじまる窓の数とする。

2 章で論じた交代再生過程の性質に基づき、阿部・鎌倉 (2016) で導出された完全な尤度関数を、2 章と同じ記号を用いて書くと以下ようになる。

$$L(\theta) = \rho^{n_1} (1 - \rho)^{n - n_1} \times \prod_{i=1}^m \left(\left[\{f_1(\tau_i)\}^{d_i} \{1 - F_1(\tau_i)\}^{1-d_i} \right]^{a_i} \left[\{g_1(\tau_i)\}^{d_i} \{G_1(\tau_i)\}^{1-d_i} \right]^{1-a_i} \right)^{s_i} \times \left(\left[\{f_0(\tau_i)\}^{d_i} \{1 - F_0(\tau_i)\}^{1-d_i} \right]^{a_i} \left[\{g_0(\tau_i)\}^{d_i} \{G_0(\tau_i)\}^{1-d_i} \right]^{1-a_i} \right)^{1-s_i}. \quad (4.1)$$

ここで θ は $F_0(\tau)$ および $F_1(\tau)$ のパラメータである。この尤度関数を最大化することでパラメータを推定することができる。多くの場合、推定量は閉じた形では求まらないため、数値的に最大化する必要がある。

5 窓打ち切りされた交代再生イベントのモデルとパラメータの推定量の比較

Rootzén (2016) では、(4.1) 式と同様の尤度関数を導出しながらも、条件付けることで状態 1 のパラメータに依存する項を削除し、条件付きの尤度関数、

$$L(\theta_0 | n_1) = \prod_{i=1}^m \left(\left[\{f_0(\tau_i)\}^{d_i} \{1 - F_0(\tau_i)\}^{1-d_i} \right]^{a_i} \left[\{g_0(\tau_i)\}^{d_i} \{G_0(\tau_i)\}^{1-d_i} \right]^{1-a_i} \right)^{1-s_i} \quad (5.1)$$

を用いた。ここで θ_0 は、 $f_0(\tau)$ に関するパラメータである。条件付き確率の定義より n_1 を所与とすると (4.1) 式の最初の $\rho^{n_1} (1 - \rho)^{n - n_1}$ の因子が消える。以降、この (5.1) 式の関数を最大化することでパラメータを推定する方法を Rootzén & Zholud 法と呼ぶ。

状態 0 の分布と状態 1 の分布がともに指数分布のとき、完全な尤度関数を用いる阿部・鎌倉 (2016) の方法の Rootzén & Zholud 法に対する漸近相対効率、

$$1 + \mu_1 \mu_0 / \{w(\mu_0 + \mu_1) + \mu_1 \mu_0\} > 1 \quad (5.2)$$

となる (阿部・鎌倉, 2017). すべてのパラメータが正であることを注意して, 漸近相対効率は 1 以上であることがわかる.

状態 0 の分布がワイブル分布の場合の推定量の性質については, 阿部・鎌倉 (2017) でシミュレーションを用いて考察されており, 小サンプルの場合や, 窓の幅に対して $F_0(\tau)$ の平均が長いとき, 阿部・鎌倉 (2016) の方法が標準誤差が小さいことが確認された. Rootzén & Zholud 法では観測が状態 1 から始まる窓の数の情報を用いていないためである.

1 章で触れたヒートシールの界面の分析は阿部・鎌倉 (2017) に詳しい. 状態 0, 状態 1 の長さの分布にそれぞれワイブル分布を仮定し, それぞれの尺度母数を $\exp(\beta_{00} + \beta_{01}x)$, $\exp(\beta_{10} + \beta_{11}x)$ と置くと, 阿部・鎌倉 (2016) の方法で回帰型の構造を持ったモデルのパラメータを推定することができる. 本研究の場合 x はヒートシール接着時の温度である. このモデルを用いることで界面の長さの分布を推定することができ, ヒートシール接着の際の適切な温度を調べることができる.

6 結論

本研究では窓打ち切りされた確率過程について考察を行った. その成果を要約する. 第一に窓打ち切りされた交代再生過程について, 尤度関数を導出し, パラメータ推定を可能にした. また, 交代再生過程における 2 つの状態の長さの分布が独立であっても, 窓打ち切り状況下では両者のパラメータを同時に推定するアプローチが, 標準誤差で評価して優れた性質を持つことを明らかにした. 条件付き尤度関数に基づく Rootzén & Zholud (2016) の方法を用いるメリットは計算が簡単になることであるが, 計算機の発達した昨今では, 全尤度関数を用いる方法もさほど大きな負担にはならない. 特殊な事情のない限り全尤度関数を用いた方法を取るべきである. さらに, 顕微鏡写真を通して観測されるヒートシールの界面を窓打ち切りされた交代再生過程として扱う新しいモデルを提案した. この成果はヒートシールのみならず, 信頼性工学におけるアベイラビリティや MTBF, MTTR の推定にも応用が可能である.

参考文献

- [1] 阿部興・鎌倉稔成 (2016). 窓打ち切り状況下での交代再生過程のパラメータの最尤推定. *計算機統計学*, **29**, 133-146.
- [2] 阿部興・鎌倉稔成 (2017). 窓打ち切りされた観測データの交代再生イベントのモデルとパラメータの推定方法および推定量の比較. *計算機統計学* (in printing).
- [3] Beichelt, F. E. & Paul Fattil, L. (2002). *Stochastic Processes and Their Applications*. Taylor & Francis.
- [4] Laslett, G. M. (1982). The survival curve under monotone density constraints with application to two-dimensional line segment processes. *Biometrika.*, **69** 153-160.
- [5] Rootzén H. & Zholud, D. (2016). Tail Estimation for Window-Censored Processes. *Technometrics.*, **58**:1, 95-103.
- [6] Zhao (2013). Parametric inference from window censored renewal process data., Ohio University.
- [7] Zhu, Y., Yashchin E. & Hosking, J. R. M. (2014) Parametric estimation for window censored recurrence data. *Technometrics.*, **56**:1, 55-66.