

中央大学大学院理工学研究科経営システム工学専攻
博士論文

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

阿部興

Ko ABE

学籍番号 14S7100002K

指導教員 鎌倉稔成 教授

目次

第 1 章	序論	7
1.1	背景	7
1.2	本論文の構成	8
第 2 章	対象とする確率過程	10
2.1	はじめに	10
2.2	対象とする量	10
2.3	ポアソン過程	11
2.4	連続時間マルコフ連鎖	13
2.5	再生過程	16
2.6	交代再生過程	21
2.7	マルコフ再生過程	24
2.8	この章のまとめ	25
第 3 章	窓打ち切り状況下での再発事象のパラメータ推定	26
3.1	はじめに	26
3.2	打ち切りや切断されたデータに対する尤度の構成	26
3.3	窓打ち切り	29
3.4	窓打ち切りされた再生過程のパラメータ推定	29
3.5	マルコフ再生過程を用いたモデル	31
3.6	窓打ち切りされた線分過程のパラメータ推定	34
3.7	この章のまとめ	42
第 4 章	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	50
4.1	はじめに	50
4.2	窓打ち切りされた交代再生過程	50
4.3	モデル	50
4.4	尤度関数の導出	51
4.5	尤度関数の左右対称性について	53

4.6	アベイラビリティの簡便な推定量	54
4.7	シミュレーションによる推定量の比較	54
4.8	この章のまとめ	56
第 5 章	窓打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	65
5.1	はじめに	65
5.2	打切りのメカニズム	65
5.3	条件付き最尤法に基づく推定	67
5.4	推定量の漸近的性質	68
5.5	シミュレーションによる推定量の比較	71
5.6	事例研究	90
5.7	この章のまとめ	91
第 6 章	まとめ	93
	参考文献	95
	付録	97
A	シミュレーションによる Laslett (1982) の推定量の検証	97
B	MTTR の推定量の比較	99

目次

2.1	ポアソン過程のサンプルパス	12
2.2	2 状態連続時間マルコフ連鎖のサンプルパス	15
2.3	余命 $Z(t)$ の模式図	18
3.1	一般化されたタイプ I 打ちりの模式図	28
3.2	窓打ちりの模式図	29
3.3	窓打ちりされた再生過程の模式図	31
3.4	観測される故障のパターンの例. ●は修理, ×は交換を表す.	33
3.5	Mauldon(1998) が扱った観測の模式図	39
4.1	窓打ちりされた交代再生過程. X_1 の従う分布と X_2 の従う分布は異なる. 窓の外側 で起きたイベントについては観測されない.	51
4.2	顕微鏡写真の模式図. 黒い線分は接着されていない界面.	51
4.3	極限アベイラビリティの推定結果. 指数分布を仮定し, パラメータ $\eta_1 = 8, \eta_2 = 2$ の とき. サンプルサイズ 10 のとき.	56
4.4	極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8,$ $\eta_2 = 2, m_1 = 3, m_2 = 3$, サンプルサイズ 10 のとき.	57
4.5	極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8,$ $\eta_2 = 2, m_1 = 5, m_2 = 5$, サンプルサイズ 10 のとき.	57
4.6	極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8,$ $\eta_2 = 2, m_1 = 7, m_2 = 7$, サンプルサイズ 10 のとき.	58
4.7	極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8,$ $\eta_2 = 2, m_1 = 0.8, m_2 = 0.8$, サンプルサイズ 10 のとき.	58
4.8	極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8,$ $\eta_2 = 2, m_1 = 0.5, m_2 = 0.5$, サンプルサイズ 10 のとき.	59
4.9	極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8,$ $\eta_2 = 2, m_1 = 0.3, m_2 = 0.3$, サンプルサイズ 10 のとき.	59
4.10	極限アベイラビリティの推定結果. 指数分布を仮定し, パラメータ $\eta_1 = 8, \eta_2 = 2,$ サンプルサイズ 5 のとき.	60

4.11	極限アベアラビリティの推定結果. 指数分布を仮定し, パラメータ $\eta_1 = 8, \eta_2 = 2,$ サンプルサイズ 100 のとき.	60
4.12	極限アベアラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8,$ $\eta_2 = 2, m_1 = 5, m_2 = 5,$ サンプルサイズ 5 のとき.	61
4.13	極限アベアラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8,$ $\eta_2 = 2, m_1 = 0.5, m_2 = 0.5,$ サンプルサイズ 5 のとき.	61
4.14	極限アベアラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8,$ $\eta_2 = 2, m_1 = 5, m_2 = 5,$ サンプルサイズ 100 のとき.	62
4.15	極限アベアラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8,$ $\eta_2 = 2, m_1 = 0.5, m_2 = 0.5,$ サンプルサイズ 100 のとき.	62
4.16	極限アベアラビリティの推定結果. ガンマ分布を仮定し, パラメータ $k_1 = 2,$ $k_2 = 2, \sigma_1 = 4, \sigma_2 = 1,$ サンプルサイズ 5 のとき.	63
4.17	極限アベアラビリティの推定結果. ガンマ分布を仮定し, パラメータ $k_1 = 2,$ $k_2 = 2, \sigma_1 = 4, \sigma_2 = 1,$ サンプルサイズ 10 のとき.	63
4.18	極限アベアラビリティの推定結果. ガンマ分布を仮定し, パラメータ $k_1 = 2,$ $k_2 = 2, \sigma_1 = 4, \sigma_2 = 1,$ サンプルサイズ 100 のとき.	64
5.1	Rootzén & Zholud (2016) が扱った 5 種類の観測.	66
5.2	μ_0 の推定値. 指数分布を仮定し, Rootzén & Zhould タイプの打切りの場合.	74
5.3	μ_0 の推定値. 指数分布を仮定し, 窓打切りの場合.	75
5.4	形状パラメータ m の推定値. ワイブル分布 ($m = 0.5$) を仮定し, Rootzén & Zhould タイプの打切りの場合.	75
5.5	尺度パラメータ η の推定値. ワイブル分布 ($m = 0.5$) を仮定し, Rootzén & Zhould タイプの打切りの場合.	76
5.6	形状パラメータ m の推定値. ワイブル分布 ($m = 2$) を仮定し, Rootzén & Zhould タイプの打切りの場合.	76
5.7	尺度パラメータ η の推定値. ワイブル分布 ($m = 2$) を仮定し, Rootzén & Zhould タイプの打切りの場合,	77
5.8	形状パラメータ m の推定値, ワイブル分布 ($m = 0.5$) を仮定し, 窓打切りの場合	78
5.9	尺度パラメータ η の推定値. ワイブル分布 ($m = 0.5$) を仮定し, 窓打切りの場合	79
5.10	形状パラメータ m の推定値. ワイブル分布 ($m = 2$) を仮定し, 窓打切りの場合.	79
5.11	尺度パラメータ η の推定値. ワイブル分布 ($m = 2$) を仮定し, 窓打切りの場合.	80
5.12	形状パラメータ k の推定値. ガンマ分布 ($k = 3$) を仮定し, Rootzén & Zhould タ イプの打切りの場合.	80
5.13	尺度パラメータ σ の推定値. ガンマ分布 ($k = 3$) を仮定し, Rootzén & Zhould タ イプの打切りの場合.	81

5.14	形状パラメータ k の推定値. ガンマ分布 ($k = 0.5$) を仮定し, Rootzén & Zhould タイプの打切りの場合.	81
5.15	尺度パラメータ σ の推定値. ガンマ分布 ($k = 0.5$) を仮定し, Rootzén & Zhould タイプの打切りの場合.	82
5.16	形状パラメータ k の推定値. ガンマ分布 ($k = 3$) を仮定し, 窓打切りの場合.	82
5.17	尺度パラメータ σ の推定値. ガンマ分布 ($k = 3$) を仮定し, 窓打切りの場合.	83
5.18	形状パラメータ k の推定値. ガンマ分布 ($k = 0.5$) を仮定し, 窓打切りの打切りの場合.	83
5.19	尺度パラメータ σ の推定値. ガンマ分布 ($k = 0.5$) を仮定し, 窓打切りの打切りの場合.	84
5.20	MTTR の推定値の比較. 指数分布を仮定し, Rootzén & Zhould タイプの打切りの場合.	85
5.21	MTTR の推定値の比較. ワイブル分布 ($m = 2$) を仮定し, Rootzén & Zhould タイプの打切りの場合.	85
5.22	MTTR の推定値の比較. ワイブル分布 ($m = 0.5$) を仮定し, Rootzén & Zhould タイプの打切りの場合.	86
5.23	MTTR の推定値の比較. ガンマ分布 ($k = 3$) を仮定し, Rootzén & Zhould タイプの打切りの場合.	86
5.24	MTTR の推定値の比較. ガンマ分布 ($k = 0.5$) を仮定し, Rootzén & Zhould タイプの打切りの場合.	87
5.25	MTTR の推定値の比較. 指数分布を仮定し, 窓打切りの場合.	87
5.26	MTTR の推定値の比較. ワイブル分布 ($m = 2$) を仮定し, 窓打切りの場合.	88
5.27	MTTR の推定値の比較. ワイブル分布 ($m = 0.5$) を仮定し, 窓打切りの場合.	88
5.28	MTTR の推定値の比較. ガンマ分布 ($k = 3$) を仮定し, 窓打切りの場合.	89
5.29	MTTR の推定値の比較. ガンマ分布 ($k = 0.5$) を仮定し, 窓打切りの場合.	89
5.30	温度と溶着率の関係. 実線: 指数分布, 点線: ワイブル分布.	92
B.1	MTTR の推定値の比較. ワイブル分布 ($m = 0.5$) を仮定し, Rootzén & Zhould タイプの打切りの場合.	100
B.2	MTTR の推定値の比較. ワイブル分布 ($m = 0.5$) を仮定し, 窓打切りの場合.	100

表目次

3.1	推定値の平均の比較. 指数分布を仮定した場合.	41
3.2	推定値の平均の比較. ワイブル分布を仮定した場合.	42
3.3	推定値の平均の比較. ガンマ分布を仮定した場合.	43
3.4	推定値の RMSE の比較. 指数分布を仮定した場合.	44
3.5	推定値の RMSE の比較. ワイブル分布を仮定した場合.	45
3.6	推定値の RMSE の比較. ガンマ分布を仮定した場合.	46
3.7	μ の被覆確率. 指数分布を仮定した場合	47
3.8	パラメータの被覆確率. ワイブル分布を仮定した場合	48
3.9	パラメータの被覆確率. ガンマ分布を仮定した場合	49
5.1	Rootzén & Zholud 法と阿部・鎌倉 (2016) の方法の分散の比較.	78
5.2	指数分布とワイブル分布のモデルにおける各パラメータの推定値と AIC.	91
A.1	ガンマ分布のパラメータの推定 ($k = 2$ のとき)	97
A.2	ガンマ分布のパラメータの推定 ($k = 0.5$ のとき)	98
A.3	ワイブル分布のパラメータの推定 ($m = 2$ のとき)	98
A.4	ワイブル分布のパラメータの推定 ($m = 0.5$ のとき)	99

第 1 章

序論

1.1 背景

本研究の動機はヒートシールの強度の分析から得られた。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、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、顕微鏡写真を用いる場合、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、界面の正確な長さそのものを知ることはできない。本論文ではこの観測の可能な区間を「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。ヒートシールの界面は線分とみなすことができる。限定された窓の視野から観測される線分の長さの分布を推定する問題は、しばしば Laslett's line segment problem と呼ばれる (Van Zwet, 2004)。本研究も Laslett's line segment problem の一端とみなすことができる。線分としてモデル化できる対象は数多く存在する。Laslett (1982) は炭鉱の壁に入るひびを線分として扱った。Mauldon (1998) は岩石のトレースを線分とし、Svensson *et al.* (2006) は樹木の繊維を線分として分析を行った。

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、打ち切りと呼ば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。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のは、実務上の制約から現実的ではない。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。一般的にはデータの打ち切りは、推定量にバイアスを生じさせる原因となる。生存時間分析が対象とするイベントは死亡とよばれる。生存時間分析が扱う対象は一度しかイベントが生起しないことを仮定している。本論文では関心のある事象が繰り返し起こる再発事象 (recurrent event) を扱う。

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、窓打ち切りされた (window-censored) という (Zhao, 2006, Zhu *et al.*, 2014, Rootzén & Zholud, 2016)。窓打ち切りという語は近年になって使われるようになったが、右打ち切りを

拡張する形で1980年代から研究がなされている。Laslett (1982) は、患者の到着に対して定常ポアソン過程を仮定して、観測開始の原点0の以前にも到着がある場合について考察を行い、生存時間分布のノンパラメトリックな最尤推定量を構成した。到着はポアソン過程であるが、生存時間は任意の分布に対応している。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。窓打切りされた再生過程のパラメータの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、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広くなされているが、交代再生過程についても研究が必要である。交代再生過程とは2つの状態が交互に生起する再生過程の一般化である。本研究の動機となったヒートシールの例では、接着面と未接着面が交互に現れる。これは交代再生過程としてモデル化できる。また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 (Barlow & Preschan, 1965)。Peuter (2013) はバスケットボールのオフェンスとディフェンスのサイクルに交代再生過程を当てはめ、勝敗を予測した。Rootzén & Zholud (2016) では自動車の運転者が道路を見ている状態を1、脇見をしている状態を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推定に基づく推定量を提案した。

1.2 本論文の構成

第2章でははじめに一般に計数過程と呼ばれるもののうち、もっとも基本的で重要な例であるポアソン過程について述べる。続けて、その一般化として再生過程、連続時間マルコフ連鎖、交代再生過程を紹介する。本論文が取り扱う交代再生過程が、多様な確率過程を含む族であることがわかるであろう。

第3章では窓打切りされた観測がこれまでどのように扱われてきたかを解説する。性質が十分に明らかにされていない Laslett (1982) および Mauldon (1998) の推定量に関してはシミュレーションによる比較検討を行う。また Laslett (1982) の方法に関しては線分の長さの分布が指数分布に従う場合、漸近分散を求めることができる。

第4章では阿部・鎌倉 (2016) に基づき、窓打切り状況下での交代再生過程のパラメータ推定を論じる。これまでに窓打切り状況下の再発事象がどのようにモデリングされてきたかを整理し、提案するパラメータの推定方法を述べる。提案する手法は最尤法に基づく。

第5章では、阿部・鎌倉 (2017) に基づき、Rootzén & Zholud (2016) の提案した条件付き最尤法に基づく推定と、阿部・鎌倉 (2016) が提案した最尤法に基づく推定の比較を行う。窓打切りされた交代再生過程については、阿部・鎌倉 (2016) による方法と Rootzén & Zholud (2016) による方法が独立に提案されたため、いまだ推定量の性質が十分に明らかになっているとは言い難い。本研究では状態0の長さの分布と、状態1の長さの分布がともに指数分布に従う場合、両者の方法の漸近分散を明らかにする。またワイブル分布、ガンマ分布の場合に関しては、シミュレーションを用いて推定量のバイアス、標準誤差を評価する。Rootzén & Zholud (2016) は状態0の長さの分布を推定することに関

心があるため、状態 1 の長さの分布のパラメータに依存する因子を条件付けて消去したのに対し、阿部・鎌倉 (2016) では、状態 1 の長さの比率に関心があり、状態 0 の長さの分布と、状態 1 の長さの分布を同時に推定するアプローチを提案している。本研究では、状態 0 の長さの分布のみに関心があるとしても、(条件付けをしない) 完全な尤度関数を用いた最尤推定量が標準誤差で比較して優れた性質を持つことを明らかにする。

また窓打ち切り状況下での回帰型のモデルについては、Zhu *et al.* (2014) による研究がある。Zhu *et al.* (2014) による提案は、マルコフ再生過程に基づくモデルである。交代再生過程における回帰モデルは阿部・鎌倉 (2016) および Rootzén & Zholud (2016) では、今後の課題とされている。実際に分析を行った例は阿部・鎌倉 (2017) にあるため、5.6 節では、それを引用し交代再生過程における回帰モデルについて論じる。

第 2 章

対象とする確率過程

2.1 はじめに

この章では、確率過程の中でも最も基本的で重要なポアソン過程から始まり、その拡張として 2 状態マルコフ過程、再生過程、交代再生過程について議論し、後の章で用いる必要な性質を紹介する。本論文は窓打切りというデータの不完全性を扱うことが主眼であるので、余命の分布と極限での振る舞いが主な関心事項である。

2.2 節では生存時間分析の分野で考察の対象となるハザード関数や生存関数と、密度関数、分布関数との関係を整理する。

ポアソン過程はイベントの生起の間隔が常に同一の指数分布に従う。それに対し、連続時間マルコフ連鎖ではイベントの生起の間隔が、状態によってそれぞれパラメータの異なる指数分布に従う。再生過程はイベントの生起の間隔が独立に同一の正の連続型の分布に従う。交代再生過程ではイベントの生起の間隔が交互に異なる分布に従う。すなわち交代再生過程は、ポアソン過程、2 状態マルコフ連鎖、再生過程の拡張と捉えることができ、多様な現象のモデリングに利用できる。

この章の内容は、Cox and Isham (1980), Ross (1995), 尾崎 (1996), Beichelt & Paul (2002) に詳しい。

2.2 対象とする量

ここでは再発事象を考えるときに用いられる基本的な量である、生存関数とハザード関数を紹介する。

確率密度関数 $f(t)$ を持つ分布 $F(t)$ に対して、 $1 - F(t)$ は生存関数 (survival function) と呼ばれる。ハザード関数 $h(t)$ は、

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

と定義される。この式は「時刻 t までイベントが生起せず、直後にイベントが発生する確率」と解釈

できる. T が連続型の確率変数のときは,

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{\Pr(t \leq T < t + \Delta t \cap T \geq t)}{\Pr(T > t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t)}{\Delta t} / \Pr(T > t) \\ &= \frac{f(t)}{1 - F(t)} \end{aligned}$$

となる.

$$H(t) = \int_0^t h(u) du$$

は累積ハザード関数と呼ばれる. 生存関数は,

$$1 - F(t) = \exp(-H(t))$$

と表すことができる. なぜならば,

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log\{1 - F(t)\}$$

より, 両辺積分し,

$$\exp\left(-\int_0^t h(u) du\right) = \{1 - F(t)\}$$

となるためである.

以上の議論より, 密度関数, 分布関数, 生存関数, ハザード関数は, いずれか1つが決まれば他を導くことができる.

2.3 ポアソン過程

ポアソン過程は確率過程の中で最も基本的で重要なものである.

$N(t)$ を区間 $[0, t)$ で起こるある事象の個数とする. 区間は時間的なものと見ても空間的なものと見ても構わない. 注目した事象が以下のような性質を持って生起するとき, $N(t)$ はポアソン過程に従うという.

1. 互いに背反する区間では, 現象の生起する確率は独立である. すなわち $s < t \leq u < v$ で $N(t) - N(s)$ と $N(v) - N(u)$ は独立である.
2. 非常に小さい区間で事象の生起する確率は, その区間の長さに比例する. すなわち $P(N(t+h) - X(t)) = \lambda h + o(h)$, ここで $o(h)$ は h より高位の無限小とよび, λ は正の定数である.
3. 非常に小さい区間で事象が2回以上生起する確率は, その区間で事象が1回起こる確率に比べ, 無視できるほど小さい. すなわち $P(N(t+h) - N(t) \geq 2) = o(h)$.
4. 区間の大きさが0ならば, 事象が生起する確率は0である. すなわち $N(0) = 0$.

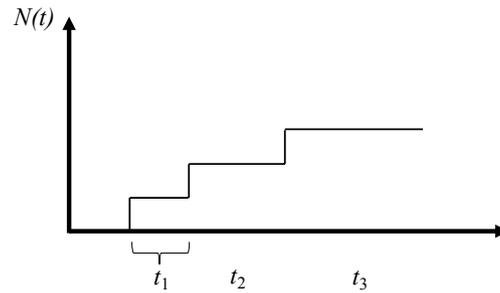


図 2.1 ポアソン過程のサンプルパス

上記の仮定のもとで、事象の生起する間隔 t_1, t_2, \dots はパラメータ λ の指数分布になる。仮定 1, 2, 3 より、

$$P(N(t+h) = 0) = (1 - \lambda h)P(N(t) = 0) + o(h)$$

が成り立つ。ここから以下の微分方程式が導かれる。

$$P'(N(t) = 0) = -\lambda P(N(t) = 0)$$

これを解くと、

$$P(N(t) = 0) = \exp(-\lambda t + C)$$

(C は積分定数) が得られる。仮定 4 より $P(N(0) = 1)$ なので、 $P(N(t) = 0) = \exp(-\lambda t)$ である。これは指数分布の生存関数と等しい。 $P(N(t) = 0)$ と $P(t_i > t)$ は同値であるから、 t_1 は指数分布に従うことが分かる。1 の独立性の仮定より、

$$\begin{aligned} P(t_2 > t | t_1 > s) &= P(N(t+s) - N(s) = 0 | N(s) = 1) \\ &= P(N(t+s) - N(s) = 0) \\ &= e^{-\lambda t} \end{aligned}$$

が成り立つ。これもまた指数分布の生存関数と等しい。

図 2.1 は t_1, t_2, \dots と $N(t)$ の関係を模式的に表したものである。

続く 2.4 節以降では、以下の性質が重要となる。

定理 1. 確率変数 X が指数分布に従うとき、

$$P(X > s+t | X > s) = P(X > t), \quad s \geq 0, t \geq 0. \quad (2.1)$$

また、この関係を満たす非負の連続型の確率変数は指数分布に従う。

まず $P(X > s+t | X > s) = P(X > s+t) / P(X > s) = P(X > t)$ は指数分布の生存関数より明らかである。次に $P(X > s+t | X > s) = P(X > t)$ が成り立つとすると、

$$P(X > s+t | X > s) = P(X > s)P(X > t).$$

$s = t = 0$ とおくと、仮定より $P(X > 0) = 1$. ここで、 $G(x) = P(X > x)$ とおくと、

$$G(s+t) = G(s)G(t).$$

これを満たすためには、 $G(t) = e^{-\lambda t}$ でなければならない.

次に、事象が発生するまでの待ち時間 X_1, X_2, \dots , が独立にパラメータ λ の指数分布に従うとき $N(t)$ はパラメータ λt のポアソン分布に従うことを示そう. $S_n = X_1 + \dots + X_n$, $S_0 = 0$ とする. $N(t) \geq n$ と $S_n \leq t$ が同値であることに注意する. まず $N(t) = 0$ となる確率を考える. $N(t) = 0$ は、最初の事象の生起までの待ち時間が、 t よりも長いということなので、

$$\Pr(N(t) = 0) = \Pr(X_1 > t) = e^{-\lambda t}$$

次に $N(t) = n (n = 1, 2, \dots)$ となる確率を考える.

$$\begin{aligned} P(N(t) = n) &= P(N(t) \geq n) - P(N(t) \geq n+1) \\ &= P(S_n \leq t) - P(S_{n+1} \leq t) \end{aligned}$$

$S_n = X_1 + \dots + X_n$ はパラメータ λ , n のガンマ分布に従うので、

$$\begin{aligned} P(N(t) = n) &= \int_0^t \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} dx - \int_0^t \frac{\lambda^{n+1} x^n}{n!} e^{-\lambda x} dx \\ &= \int_0^t \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} dx - \frac{\lambda^{n+1} x^n}{n!} e^{-\lambda x} dx \\ &= \left[\frac{(\lambda x)^n}{n!} e^{-\lambda x} \right]_0^t \\ &= \frac{(\lambda t)^n}{n!} e^{-\lambda t} \end{aligned}$$

これはパラメータ λt のポアソン分布の確率関数である.

2.4 連続時間マルコフ連鎖

ポアソン過程ではイベントの生起の間隔は常に同一の指数分布に従う. 連続時間マルコフ連鎖は、複数の異なる状態を持ち、状態によって再生間隔がパラメータの異なる指数分布に従う.

非負の整数の値を取る連続時間の確率過程 $\{X(t), t \geq 0\}$ を考える. 任意の $s, t \geq 0$ および非負の整数 $i, j, x(u), 0 \leq u \leq s$ に対し、

$$\Pr\{X(t+s) = j | X(s) = i, X(u) = x(u), 0 \leq u < s\} = \Pr\{X(t+s) = j | X(s) = i\}.$$

が成り立つとき、この過程を連続時間マルコフ連鎖 (continuous-time Markov chain) と呼ぶ. すなわち、連続時間マルコフ連鎖は、マルコフ性を持つ. s での状態が与えられたときの、 $t+s$ における状態は現在の状態のみに依存し、過去には依存しない.

加えて、 $P\{X(t+s) = j | X(s) = i\}$ が s と独立であるとき、定常 (stationary) または斉次的 (homogeneous) な推移確率を持つという. 以下では定常な推移確率を持つ連続時間マルコフ連鎖のみ

を考える。マルコフ性により、区間 $[s, s+t]$ で同じ状態に留まる確率は、その状態に少なくとも時間 t 留まる確率と一致する。すなわち過程が状態 i に留まる時間を τ_i と表記すると、任意の $s, t \geq 0$ に対し以下が成立する。

$$\Pr\{\tau_i > s+t | \tau_i > s\} = \Pr(\tau_i > t).$$

従い τ_i は無記憶性を持つため、必ず指数分布に従う。

以下の式で表される t 時間後に過程が i から j になる確率を考える。

$$P_{ij}(t) = P\{X(t+s) = j | X(s) = i\}. \quad (2.2)$$

以降 $P_{ij}(t)$ を求めるためにマルコフ性を利用して、微分方程式を導くが、そのために以下で定義される推移率を導入する。

$$v_i = \lim_{\Delta t \rightarrow 0} \frac{1 - P_{ii}(\Delta t)}{\Delta t},$$

$$q_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(\Delta t)}{\Delta t}, \quad i \neq j.$$

$$P_{ij}(t+h) = \sum_{k=0}^{\infty} P_{ik}(t) P_{kj}(h)$$

より、

$$\begin{aligned} P_{ij}(t+h) - P_{ij}(t) &= \sum_k P_{ik}(t) P_{kj}(h) - P_{ij}(t) \\ &= \sum_{k \neq j} P_{ik}(t) P_{kj}(h) - \{1 - P_{jj}(h)\} P_{ij}(t) \end{aligned}$$

従い、

$$\lim_{h \rightarrow 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h} = \lim_{h \rightarrow 0} \left\{ \sum_{k \neq j} P_{ik}(t) \frac{P_{kj}(h)}{h} - \frac{1 - P_{jj}(h)}{h} P_{ij}(t) \right\}$$

和と極限の順序交換が可能であると仮定して、

$$P'_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t) \quad (2.3)$$

式 (2.3) はコルモゴロフの前向き方程式 (Kolmogorov's forward equations) と呼ばれる。

コルモゴロフの前向き方程式を解くことにより、 $P_{ij}(t)$ を求めることができる。応用上特に重要な2状態の連続時間マルコフ連鎖について、 $P_{ij}(t)$ を求めることにする。過程のとり状態が0と1の2種類とすると、以下の前向き方程式を得る。

$$\begin{aligned} P'_{00}(t) &= \mu P_{01}(t) - \lambda P_{00}(t) \\ &= -(\lambda + \mu) P_{00}(t) + \mu, \end{aligned}$$

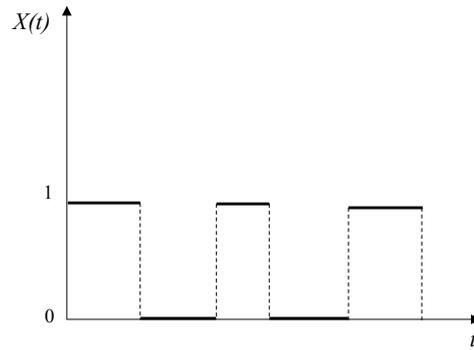


図 2.2 2 状態連続時間マルコフ連鎖のサンプルパス

ここで2つめの等号は $P_{01}(t) = 1 - P_{00}(t)$ より成り立つ. 右辺第1項を移行し, 両辺に $e^{(\lambda+\mu)t}$ を掛けると, 以下の等式が導かれる.

$$e^{(\lambda+\mu)t}[P'_{00}(t) - (\lambda + \mu)P_{00}(t)] = \mu e^{(\lambda+\mu)t}$$

これより,

$$\frac{d}{dt}[e^{(\lambda+\mu)t}P'_{00}(t)] = \mu e^{(\lambda+\mu)t}$$

であるから, 両辺を積分することにより以下を得る.

$$e^{(\lambda+\mu)t}P'_{00}(t) = \frac{\mu}{\lambda + \mu}e^{(\lambda+\mu)t} + C,$$

ここで C は積分定数. 過程が状態 0 から始まるとすると, $P_{00}(t) = 1$ であるから, $C = \lambda/(\lambda + \mu)$ に決まる. 結果, 以下のように $P_{00}(t)$ が求まる.

$$P'_{00}(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu}e^{-(\lambda+\mu)t}.$$

同様の方法で $P_{11}(t)$ は以下のように求まる.

$$P'_{11}(t) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu}e^{-(\lambda+\mu)t}.$$

2 状態連続時間マルコフ連鎖は稼働と休止を繰り返すシステムのモデルとして用いることができる. その場合状態 0 を休止, 状態 1 を稼働に対応させる. この 2 状態連続時間マルコフ連鎖のサンプルパスを図 2.2 に示す.

十分に時間が経過した後, システムが稼働している確率は以下のように求まる.

$$\lim_{t \rightarrow \infty} P'_{11}(t) = \frac{\lambda}{\lambda + \mu}. \quad (2.4)$$

この量は, 信頼性工学の分野ではアベイラビリティと呼ばれ, システムの可用性の指標となる. アベイラビリティの推定方法については 4 章で述べる.

2.5 再生過程

2.3 節で述べ通り, ポアソン過程では, 事象の生起する間隔は指数分布に従う. その自然な一般化として, 事象の生起する間隔が, 指数分布とは限らない一般の分布に従う確率過程を考えると, これを再生過程と呼ぶ. 本節ではその基本的な性質を整理する.

$\{X_i, i = 1, 2, \dots, \}$ を同一の分布 $F(x)$ に従う非負の独立な確率変数の列とする. $S_0 = 0$, $S_n = \sum_{i=1}^n X_i$ とし,

$$N(t) = \sup\{n | S_n \leq t\} \quad (2.5)$$

とする. $N(t)$ は t までに生起したイベントの個数である. 計数過程 $\{N(t), t \geq 0\}$ を再生過程 (renewal process) と呼ぶ. $N(t)$ の平均 $M(t) = E[N(t)]$ は再生関数 (renewal function) と呼ばれる. S_n は独立で同一の分布 $F(t)$ にしたがう確率変数の和であるから, S_n の分布は, $F(x)$ の n 重たたみこみ ($F^{(n)}(t)$ と表記する) によって求まる. 再生関数 $M(t)$ は,

$$M(t) = \sum_{n=1}^{\infty} F^{(n)}(t)$$

である. なぜならば, $I_{\{n,t\}}$ を n 回目の再生が区間 $[0, t]$ 内で起こったときに 1, さもなくば 0 の値を取る変数とすると, 以下が成り立つためである.

$$\begin{aligned} E[N(t)] &= E \left[\sum_{n=1}^{\infty} I_{\{n,t\}} \right] \\ &= \sum_{n=1}^{\infty} E[I_{\{n,t\}}] \\ &= \sum_{n=1}^{\infty} \Pr\{I_{\{n,t\}} = 1\} \\ &= \sum_{n=1}^{\infty} \Pr\{S_n \leq t\} \\ &= \sum_{n=1}^{\infty} F^{(n)}(t). \end{aligned}$$

さらに再生関数 $M(t)$ は,

$$\begin{aligned}
 M(t) &= \sum_{n=1}^{\infty} F^{(n)}(t) \\
 &= F(t) + \sum_{n=1}^{\infty} F^{(n+1)}(t) \\
 &= F(t) + \sum_{n=1}^{\infty} F * F^{(n)}(t) \\
 &= F(t) + F * \left[\sum_{n=1}^{\infty} F^{(n)}(t) \right] \\
 &= F(t) + F * M(t)
 \end{aligned}$$

とも書ける. ここで $F * G(t)$ は $F(t)$ と $G(t)$ のたたみこみを表す. 積分表示すると以下のようになる.

$$M(t) = F(t) + \int_0^t M(t-x) dF(x) \quad (2.6)$$

式 (2.6) は再生方程式 (renewal equation) と呼ばれる. 再生方程式は分布のたたみこみを含むため, ラプラス変換によって簡潔になる. 一般に $F(t)$ のラプラス変換を

$$F^*(s) = \int_0^{\infty} e^{-st} dF(t) \quad (2.7)$$

と表すことにする. $M(t)$ のラプラス変換は,

$$M^*(s) = \sum_{n=1}^{\infty} [F^*(s)]^n$$

となる. これは無限等比級数の和であるので, 結果として以下の関係を得る.

$$M^*(s) = \frac{F^*(s)}{1 - F^*(s)}$$

$F^*(s)$ はモーメント母関数であるから, テイラー展開より $F^*(s) = 1 - \mu s + o(s)$ であるから,

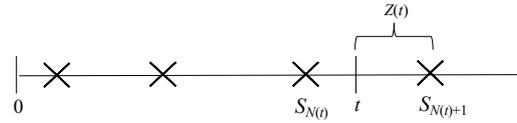
$$\lim_{s \rightarrow 0} sM^*(s) = \frac{F^*(s)}{\{1 - F^*(s)\}/s} = \frac{1}{\mu} \quad (2.8)$$

となる.

ラプラス変換について, 以下の定理が知られている (Feller, 1971; XIII 章 5 節).

定理 2. $F(t)$ が台 $[0, \infty)$ を持つ正の連続型の分布とする. $F(t)$ がラプラス変換 $F^*(t)$ を持つとする. ある非負の数 α に対して,

$$\lim_{s \rightarrow +0} s^\alpha F^*(s) = C \quad (2.9)$$

図 2.3 余命 $Z(t)$ の模式図

ならば,

$$\lim_{t \rightarrow \infty} \frac{F(t)}{t^\alpha} = \frac{C}{\Gamma(\alpha + 1)} \quad (2.10)$$

となる.

定理 2 および (2.8) 式より,

$$\lim_{t \rightarrow \infty} \frac{M(t)}{t} = \frac{1}{\mu} \quad (2.11)$$

となる. (2.11) 式の関係は基本再生定理 (elementary renewal theorem) と呼ばれる.

さらに, 到着時間間隔 X_k が連続形ならば, (2.11) 式の一般化として, 以下の関係が知られている (Ross, 1995; p.110).

$$\lim_{t \rightarrow \infty} M(t + \tau) - M(t) = \frac{\tau}{\mu} \quad (2.12)$$

(2.12) 式はブラックウェルの定理 (Blackwell's theorem) と呼ばれる. ブラックウェルの定理より,

$$\lim_{t \rightarrow \infty} \frac{M(t + \tau) - M(t)}{\tau} = \frac{1}{\mu}$$

となる. 極限操作が交換可能であるとして, 以下の関係を得る.

$$\lim_{t \rightarrow \infty} \frac{dM(t)}{dt} = \frac{1}{\mu} \quad (2.13)$$

$m(t) = dM(t)/dt$ は再生密度 (renewal density) と呼ばれ, その時刻 t における再生の起こる割合を表す.

2.5.1 再生過程における余命の分布

打切りデータを扱うにあたって, 余命 (residual life) の分布が重要になる. ここで,

$$Z(t) = S_{N(t)+1} - t \quad (2.14)$$

とおくと, $Z(t)$ は時刻 t 以降に再生が起きるまでの時間である (図 2.3).

Z_t の生存関数は、以下のように表せる。

$$\Pr(Z(t) > x) = 1 - F(t+x) + \int_0^t m(t-u)[1 - F(u+x)] du \quad (2.15)$$

ここで第1項は t 以降の最初の再生が再生過程の最初の点である確率であり、第2項は t 以前に生じた最後の点が $t-u$ にあり、続く次の点が $u+x$ 以上である確率を表す。ここで $u \in [0, t)$ である。時間が十分経過した時刻 t 以降の再生過程を考えると、以下が得られる。

$$\lim_{t \rightarrow \infty} \Pr(Z(t) > x) = \frac{\int_x^\infty [1 - F(u)] du}{\mu}. \quad (2.16)$$

(2.16) 式で表される生存関数を Ross(1996) に習い、均衡分布 (equilibrium distribution) の生存関数と呼ぶことにする。

2.5.2 均衡分布の具体的な例

指数分布の均衡分布

指数分布の均衡分布は簡単に求まる。平均 μ を持つ指数分布の分布関数は以下のように表せる。

$$F(x) = 1 - \exp\left(-\frac{x}{\mu}\right). \quad (2.17)$$

均衡分布の生存関数は、

$$G(x) = \int_x^\infty \frac{1}{\mu} \exp\left(-\frac{u}{\mu}\right) du \quad (2.18)$$

$$= \exp\left(-\frac{x}{\mu}\right) \quad (2.19)$$

となり、元の指数分布の生存関数と一致する。

さらに逆も成り立つ。すなわち、均衡分布が元の分布と一致する確率分布は、指数分布に限られる。ここで、 $F(x)$ を台 $(0, \infty)$ を持つ分布関数とする。分布 F の平均を μ とし、 $G(x) = \int_x^\infty \frac{1 - F(u)}{\mu} du$ とする。このとき、 $1 - F(x) = G(x)$ とすると、 $F(x) = 1 - e^{-\frac{x}{\mu}}$ (指数分布) である。

$G(x)$ を微分し、

$$\frac{d}{dx} G(x) = -\frac{1 - F(x)}{\mu}.$$

$1 - F(x) = G(x)$ より、

$$\frac{d}{dx} G(x) = -\frac{1}{\mu} G(x).$$

これは1階線形微分方程式であり、以下のように解が求まる。

$$G(x) = e^{-\frac{1}{\mu}x} + C,$$

ここで C は積分定数である。仮定より $G(0) = 1 - F(0) = 1$ であるから、 $C = 0$ となり、 $F(x) = 1 - e^{-\frac{x}{\mu}}$ が得られた。また、微分方程式の解の一意性より $1 - F(x) = G(x)$ が成立するのは、指数分布の場合に限られることが分かる。

ワイブル分布の均衡分布

ワイブル分布の均衡分布の生存関数を求める。ワイブル分布の分布関数は以下の式で表される。

$$F(x) = 1 - \exp \left\{ - \left(\frac{x}{\eta} \right)^m \right\}. \quad (2.20)$$

ワイブル分布の平均は $\eta\Gamma(1 + 1/m)$ であるから、均衡分布の生存関数は以下ようになる。

$$G(x) = \int_x^\infty g(u) du = \frac{1}{\eta\Gamma(1 + 1/m)} \int_x^\infty \exp \left\{ - \left(\frac{u}{\eta} \right)^m \right\} du. \quad (2.21)$$

ここで、

$$v = \left(\frac{u}{\eta} \right)^m$$

とおくと、

$$\int_x^\infty \exp \left\{ - \left(\frac{u}{\eta} \right)^m \right\} du = \frac{1}{m} \eta \Gamma(1/m, \left(\frac{x}{\eta} \right)^m)$$

であるから、均衡分布の生存関数は以下ようになる。

$$G(x) = \frac{\Gamma \left(\frac{1}{m}, \left(\frac{x}{\eta} \right)^m \right)}{\Gamma \left(\frac{1}{m} \right)}. \quad (2.22)$$

ガンマ分布の均衡分布

次にガンマ分布の均衡分布の生存関数を求める。ガンマ分布の分布関数は以下の式で表される。

$$F(x) = \frac{\gamma(k, x/\sigma)}{\Gamma(k)} \quad (2.23)$$

ガンマ分布の平均は $k\sigma$ であるから、

$$g(x) = \frac{\Gamma(k, x/\sigma)}{\Gamma(k+1)\sigma} \quad (2.24)$$

ラプラス変換を用いると計算が容易になり、均衡分布の生存関数は以下ようになる。

$$\begin{aligned} G(x) &= \int_x^\infty g(u) du = \int_x^\infty \frac{\Gamma(k, u/\sigma)}{\Gamma(k)k\sigma} du \\ &= \frac{1}{\Gamma(k+1)} \left(\Gamma(k+1, x/\sigma) - \frac{x}{\sigma} \Gamma(k, x/\sigma) \right) \end{aligned} \quad (2.25)$$

2.6 交代再生過程

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. 2.4 節で論じた2状態マルコフ過程では, イベントの生起の間隔は指数分布であったがここでは一般の分布を考える. すなわち交代再生過程は2状態マルコフ過程の一般化と見ることできる.

互いに相異なる正の連続型の確率分布関数, F_1, F_0 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_0, μ_1 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. t 時点におけるプロセスでの状態を表すために,

$$V(t) = I\{S_{N(t)-1} + X_{N(t)} > t\} \quad (2.26)$$

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.

交代再生過程においても2状態マルコフ過程と同様に以下が成り立つ.

定理 3. 交代再生では状態1にいる確率の極限は,

$$\lim_{t \rightarrow \infty} \Pr\{V(t) = 1\} = \mu_1 / (\mu_0 + \mu_1). \quad (2.27)$$

定理3を示す前に以下の事実を確認しておく.

$$\begin{aligned} \Pr(S_{N(t)} \leq s) &= \sum_{n=0}^{\infty} \Pr(S_n \leq s, S_{N+1} > t) \\ &= 1 - F(t) + \sum_{n=1}^{\infty} \Pr(S_n \leq s, S_{N+1} > t) \\ &= 1 - F(t) + \sum_{n=1}^{\infty} \int_0^{\infty} P(S_n \leq x, S_{n+1} > t | S_n = y) dF_n(y) \\ &= 1 - F(t) + \sum_{n=1}^{\infty} \int_0^{\infty} \{1 - F(t - y)\} dF_n(y) \\ &= 1 - F(t) + \int_0^{\infty} \{1 - F(t - y)\} d\left(\sum_{n=1}^{\infty} F_n(y)\right) \\ &= 1 - F(t) + \int_0^s \{1 - F(t - y)\} dm(y). \end{aligned} \quad (2.28)$$

前回の再生の時点を条件として,

$$P(t) = \Pr(V(t) = 1 | S_{N(t)=0})P(S_{N(t)=0}) + \int_0^\infty \Pr\{V(t) = 1 | S_{N(t)=y}\} dF_{S_{N(t)}}(y). \quad (2.29)$$

いま,

$$\Pr(V(t) = 1 | S_{N(t)=0}) = P(Z_1 > t | Z_1 + Y_1 > t) \quad (2.30)$$

$$= \frac{1 - F_Z(t)}{1 - F_{Z+Y}(t)}, \quad (2.31)$$

さらに $0 < y < t$ のとき,

$$P(t) = \Pr\{V(t) = 1 | S_{N(t)} = y\} = \Pr\{Z > t - y | Z + Y > t - y\} \quad (2.32)$$

従い, (2.28) の関係を用いて,

$$P(t) = 1 - F_Z(t) + \int_0^t (1 - F_Z(t)) dM(y), \quad (2.33)$$

ここで $M(y) = \sum_{n=1}^\infty F_n(y)$. いま, $1 - F_Z(t)$ は非負, 非増加で, $\int_0^\infty (1 - F_Z(t)) dt = E[Z] < \infty$ であり, $t \rightarrow \infty$ のとき $1 - F_Z(t) \rightarrow 0$. (2.13) 式の関係を用いて,

$$\lim_{t \rightarrow \infty} P(t) = \frac{\lim_{t \rightarrow \infty} \int_0^t 1 - F_Z(t) dt}{\mu_{Z+Y}} = \frac{E[Z_n]}{E[Z_n] + E[Y_n]}, \quad (2.34)$$

2.6.1 交代再生過程における余命の分布

交代再生過程において,

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

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

と置く ($n = 1, 2, \dots$). ただし $S_{11} = X_1$ とする. また,

$$F_{S_{1n}}(t) = \Pr(S_{1n} \leq t)$$

$$F_{S_{0n}}(t) = \Pr(S_{0n} \leq t)$$

と置く. $N_a(t)$ と $N_e(t)$ をそれぞれ区間 $(0, t]$ 内で生起する状態 1 の回数と状態 0 の回数を表す変数とする. 再生関数のときと同様に期待値を取り,

$$M_1(t) = E[N_a(t)]$$

$$M_0(t) = E[N_e(t)]$$

と置く. これらは以下のように求まる.

$$M_1(t) = \sum_{n=1}^{\infty} F_1 * (F_1 * F_0)^{(n-1)}(t)$$

$$M_0(t) = \sum_{n=1}^{\infty} (F_1 * F_0)^{(n)}(t).$$

$Z_1(t)$ を過程が t 時点で状態 1 にあるときの余命を表す変数とする.

$$\Pr(V(t) = 1, Z_1(t) > x) = \sum_{n=0}^{\infty} \Pr(S_{0n} < t, S_{0n} + X_{n+1} > t + x)$$

$$= 1 - F_1(t + x) + \int_0^t \{1 - F_1(u + x)\} dM_0(t - u)$$

ここで第 1 項は t 以降の最初の再生が再生過程の最初の点である確率であり, 第 2 項は t 以前に生じた最後の点が $t - u$ にあり, 続く次の点が $u + x$ 以上である確率を表す. 時間が十分経過した時刻 t 以降のイベントの生起を考えると, 以下が得られる.

$$\lim_{t \rightarrow \infty} \Pr(V(t) = 1, Z_1(t) > x) = \frac{1}{E[X] + E[Y]} \int_x^{\infty} \{1 - F_1(u)\} du. \quad (2.35)$$

(2.34) 式より, $\lim_{t \rightarrow \infty} \Pr(V(t) = 1) = E[X] / \{E[X] + E[Y]\}$ であるから,

$$\lim_{t \rightarrow \infty} \Pr(Z_1(t) > x | V(t) = 1) = \frac{\int_x^{\infty} \{1 - F_1(u)\} du}{E[X_n]}. \quad (2.36)$$

過程が t 時点で状態 0 にあるときの余命も同様の議論により求まり, $Z_0(t)$ 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.

$$\lim_{t \rightarrow \infty} \Pr(Z_0(t) > x | V(t) = 0) = \frac{\int_x^{\infty} \{1 - F_0(u)\} du}{E[Y_n]}. \quad (2.37)$$

すなわち交代再生過程においても, 再生過程と同様, 余命の分布は均衡分布に従うことが分かる.

2.6.2 信頼性工学における交代再生過程

交代再生過程は稼働と休止を定期的に繰り返すシステムの信頼性のモデルとして用いられる. その場合, 状態 1 はシステムの稼働, 状態 0 はシステムの休止にそれぞれ対応する. またシステムが稼働している時間間隔の期待値 $E[X_i]$ は MTBF (平均故障間隔, mean time between failure の略), システムが休止している時間間隔の期待値 $E[Y_i]$ は MTTR (平均修復時間, mean time to repair の略) と呼ばれる.

また交代再生過程の状態それぞれの再生間隔の分布とともに, プロセスがある時刻 t に状態 1 にいる割合にも興味がある. この量はアベイラビリティとよばれ, 信頼性工学の分野ではシステムの可用性の指標となる.

これらの分野では、定常状態に達する前のアベイラビリティを計算することに関心を置いた研究が多い。Baxter & Li (1994) は経験分布関数に基づき、点アベイラビリティ、すなわち定常状態に達する前のアベイラビリティの推定量を構成した。Pham-Gia & Turkkan(1999) は再生間隔がガンマ分布に従う交代再生過程について、パラメータ推定に対する考察も行った。これは完全な観測が得られることを前提とした分析である。欠測データが含まれる場合の研究としては、Crow (1988) のワイブル過程を用いた信頼度成長のモデリングが挙げられる。ワイブル過程とは故障強度関数をワイブル分布のハザードとした非定常ポアソン過程である。非定常ポアソン過程はシステムの休止は考慮せず、再発する故障イベントやソフトウェアのバグの発見数をモデル化するために用いられる。

2.7 マルコフ再生過程

マルコフ再生過程は直前の状態に依存して状態が変化するが、状態の変化の間隔の分布が必ずしも指数分布には従わない。時間 t における状態を $Z(t)$ と表記する。状態 i に続く状態が j である確率を $P_{ij}(i, j \geq 1)$ とする。状態 i に続く状態が j であるという条件の下で、 i から j に遷移するまでの時間は分布 $F_i(t)$ に従う。 $F_i(t)$ の平均を μ_i と表記する。平均 μ_i が存在すること、 $F_i(t)$ が正の連続型の分布であることは仮定する。

上記の仮定の下で、

$$P_i = \lim_{t \rightarrow \infty} \Pr\{Z(t) = i | Z(0) = j\}$$

が存在し、

$$P_i = \frac{\mu_i}{\sum_j \mu_j}$$

となる。このことは状態 i を「稼働」、 i 以外の状態を「休止」とみなすことにより、交代再生過程における式 (2.34) の関係より直ちにわかる。

ここで $Y_i(j)$ を j 回目の訪問で状態 i に滞在した時間とする。 $N_i(m)$ を最初の m 回の状態遷移で、状態 i に訪れた回数とする、最初の m 回の状態遷移で状態 i に時間の割合を $P_{i=m}$ と表すと、以下のようになる。

$$\begin{aligned} P_{i=m} &= \frac{\sum_{j=1}^{N_i(m)} Y_i(j)}{\sum_i \sum_{j=1}^{N_i(m)} Y_i(j)} \\ &= \frac{\frac{N_i(m)}{m} \sum_{j=1}^{N_i(m)} \frac{Y_i(j)}{N_i(m)}}{\sum_i \frac{N_i(m)}{m} \sum_{j=1}^{N_i(m)} \frac{Y_i(j)}{N_i(m)}}. \end{aligned}$$

$m \rightarrow \infty$ として $N_i(m) \rightarrow \infty$ のとき、大数の法則より

$$\sum_{j=1}^{N_i(m)} \frac{Y_i(j)}{N_i(m)} \rightarrow \mu_i,$$

である。また、以下のマルコフ連鎖の定常確率が存在することを仮定する。

$$\lim_{m \rightarrow \infty} \frac{N_i(m)}{m} = \pi_i.$$

このとき、以下が成立する

$$\lim_{m \rightarrow \infty} P_{i=m} = \frac{\pi_i \mu_i}{\sum_j \pi_j \mu_j}. \quad (2.38)$$

2.8 この章のまとめ

本章では分析の対象となる確率過程の定義を述べ、それらの性質、主に過程が十分に長い時間が経過した極限でどのような振る舞いをするかを整理した。ここで扱った均衡分布は、続く3-5章で窓打ちデータを扱う際に重要な役割を果たす。以降の章では十分に長いプロセスを部分的に観測することを仮定するため、観測される最初のイベントの生起の間隔は均衡分布に従う。尤度関数を導くためには、均衡分布が求まる必要がある。また定理3からは、交代再生過程がある時点で状態1にいるか、状態0にいるかという点も、再生間隔の分布の平均に影響されることが分かる。種々の期待値は推定量の性質を調べるために有用である。

第3章

窓打切り状況下での再発事象のパラメータ推定

3.1 はじめに

この章では窓打切り状況下における推定量がこれまでどのように構成されてきたのかを述べる。続く4-5章で述べる窓打切り状況下における交代再生過程のパラメータの最尤推定が、本論文の主結果であるが、それらは Vardi (1982), Laslett (1982) の結果を踏まえたものである。

窓打切りについて議論する前に、様々な打切りや切断についてまとめておくことは有益である。(3.2) 節では、打切りや切断されたデータに基づき最尤推定を行う方法を述べる。これについては Klein and Moeschberger (2003) に詳しい。

3.4 節および 3.5 では窓打切り状況下での再生過程、マルコフ再生過程について、尤度関数がどのように構成されるかを述べる。

3.6 節では線分過程と呼ばれる空間上のランダムな線分の集合について議論する。Laslett(1982) は線分の長さの分布を推定する問題と生存時間の長さの分布を推定する問題の類似性を指摘し、生存時間分析のアプローチで推定量を導出した。それに対し、Mauldon (1998) はより直観的なアプローチで、線分の長さや数の平均の推定量を提案している。Laslett (1982) は尤度関数を導出したが、推定量の性質に関しては論じていない。シミュレーションによる Laslett (1982) の推定量の評価は阿部・鎌倉 (2014) で行われた。3.6 節では Mauldon(1998) と Laslett (1982) の推定量について、改めてシミュレーションによる比較を行う。

3.2 打切りや切断されたデータに対する尤度の構成

生存時間分析はイベント発生までの時間についてのデータを解析する。このようなデータは、しばしば打切りや切断といった、不完全性を伴う。

はじめに混同されがちな用語について整理を行う。その上で、打切りまたは切断されたデータに基づき構成される尤度関数がどのようになるか述べる。

右打ち切りはイベントの生起が打ち切り時間 C_r よりも長いことをいう。左打ち切りという語には注意が必要である。Turnbull(1974)では対象者の生存時間が打ち切り時間 C_l よりも小さいとき、生存時間が左打ち切りされたという。一方、Rootén and Zholud (2016)では、再生間隔が完全には得られず、ある時点からの余命として観測される場合を、左打ち切り (left-censored) と呼んでいる。本論文では、Rootén and Zholud (2016)の意味での左打ち切りを単に左打ち切りと呼び、Turnbull(1974)の意味での左打ち切りに関して触れる場合は、その都度明記することにする。

区間打ち切りは、イベント発生がある区間 $[L, R]$ に起こったことが認知され、イベントが発生した完全な時点はわからないことを指す。

打ち切り (censored) と切断 (truncated) は明確に区別されなければならない。打ち切りでは少なくとも打ち切られた被験者の数は認知されるが、切断では認知されない。一定時間 (Y_L) 以上の生存時間を持つ集団のみが観測される場合、それを左切断と呼ぶ。反対に一定時間 (Y_R) 以下の生存時間を持つ集団のみが観測される場合、それを右切断と呼ぶ。イベントの生起がある区間 $[Y_L, Y_R]$ 内に発生した場合に限り観測がされる場合、これを区間切断と呼ぶ。

これら種々の打ち切り・切断データから尤度を構成するには以下の因子を掛け合わせればよい。

- 完全に観測された生存時間. $f(x)$.
- 右打ち切りされた観測. $S(C_r)$.
- Turnbull(1974)の意味で左打ち切りされた観測. $1 - S(C_l)$.
- 区間打ち切りされた観測. $S(L) - S(R)$.
- 左切断された観測. $f(x)/S(Y_L)$.
- 右切断された観測. $f(x)/F(Y_R)$.
- 区間切断された観測. $f(x)/[S(Y_L) - S(Y_R)]$

一般化されたタイプ I 打ち切り

以降の節で扱う窓打ち切りは一般化されたタイプ I 打ち切りの拡張であるため、一般化されたタイプ I 打ち切りについて論じる。

一般化されたタイプ I 打ち切りは各対象者が異なった時点で調査研究対象者として登録され、調査研究終了時点が定められている場合に生じる (図 3.1)。

観測に一般化されたタイプ I 打ち切りが生じる場合の尤度を構成することを考える。右側打ち切りを含む生存時間データは組の変数 (Z_i, d_i) で表すことができる。 d_i は生存時間 X_i が完全に観測されたとき 1 さもなくば 0 の値を取る。患者の登録時間が生存時間と独立であるとする、打ち切り時間 C_i は X_i と独立である。 T_i は生存時間が完全に観測されたとき X_i に等しく、さもなくば C_i に等しい。すなわち $Z_i = \min(X_i, C_i)$ である。 C_i の確率密度関数を $f_C(x)$ の分布関数を $F_C(x)$ で表すと、観測

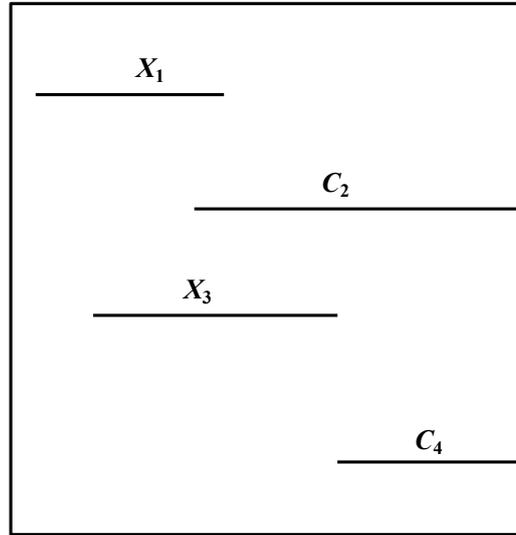


図 3.1 一般化されたタイプ I 打切りの模式図

$(Z_i = z_i, d_i = 0)$ 得られる確率は、以下のように書ける.

$$\begin{aligned}
 \Pr(Z_i = z_i, d_i = 0) &= \Pr(C_i = z_i, X_i > C_i) \\
 &= \frac{d}{dz_i} \int_0^{z_i} \int_v^{\infty} f(u) f_C(v) \, dudv \\
 &= \frac{d}{dz_i} \int_0^{z_i} \{1 - F(v)\} f_C(v) \, dv \\
 &= \{1 - F(z_i)\} f_C(z_i).
 \end{aligned}$$

同様に $d_i = 1$ のときを考えると,

$$\begin{aligned}
 \Pr(Z_i = z_i, d_i = 1) &= \Pr(X_i = z_i, X_i < C_i) \\
 &= \frac{d}{dz_i} \int_0^{z_i} \int_v^{\infty} f(u) f_C(v) \, dudv \\
 &= \frac{d}{dz_i} \int_0^{z_i} f(v) \{1 - F_C(v)\} \, dv \\
 &= f(z_i) \{1 - F_C(z_i)\}.
 \end{aligned}$$

よってこの打切りに基づく尤度を構成すると以下ようになる.

$$L = \prod_{i=1}^n [f(z_i) \{1 - F_C(z_i)\}]^{d_i} [1 - F(z_i) \{f_C(z_i)\}]^{1-d_i}$$

$F(x)$ に関するパラメータは $F_C(x)$ に関するパラメータには依存しないため、生存時間について調べるためには以下の尤度を用いればよい.

$$L = \prod_{i=1}^n [f(z_i)]^{d_i} [1 - F(z_i)]^{1-d_i}.$$

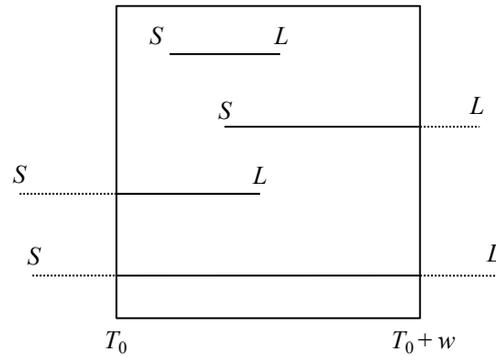


図 3.2 窓打切りの模式図

生存時間分析の分野では非常に良く知られた Kaplan-Meier 推定量は、この尤度を最大化するノンパラメトリックな推定量と見ることができる (Andersen *et al.*, 1995).

3.3 窓打切り

ある時点 S から関心のあるイベントの生起までの間隔 X を考える. イベントは $L = S + X$ で時点で生起する. S が観測されず観測開始の原点 T_0 ($T_0 > S$) からのイベントの生起までの間隔 $L - T_0$ が得られる場合, (Rootén and Zholud, 2016 の意味で) 左打切りが生じたという.

窓打切りでは右打切りと左打切りが混在する.

観測範囲が区間 $[T_0, T_0 + w]$ ($w > 0$) に限られ, この区間で生起したイベントのみが観測されるとき, その観測のスキームを窓打切りと呼ぶ. 図 3.2 は窓打切りの模式図である. 一般化されたタイプ I 打切りの模式図と比較されたい. 生存時間分析の文脈で言えば, 一般化されたタイプ I 打切りに加え, 観測開始の原点の前にエントリーされた患者の生存時間も観測される場合, 窓打切りが生じる.

3.2 節に述べた区間打切りや, 左切断と窓打切りが異なるものであることに注意が必要である.

3.4 窓打切りされた再生過程のパラメータ推定

3.4.1 窓打切りされた再生過程

窓打切りされた再生過程は, 確率過程の文献では遅れのある再生過程 (delayed renewal process) と呼ばれるものの一つである. $\{X_n, n = 1, 2, \dots\}$ を独立な非負の確率変数とする. X_1 は分布 $H(x)$ に従い, X_n ($n > 1$) は分布 $F(x)$ に従う. $S_0 = 0$, $S_n = \sum_{i=1}^n X_i$ とし, $N(t) = \sup\{n : S_n \leq t\}$ とする. 確率過程 $\{N(t), t \geq 0\}$ を遅れのある再生過程 (delayed renewal process) と呼ぶ (Ross, 1996).

以降では, $H(x)$ が 2.5.1 で述べた均衡分布である場合を考える.

3.4.2 Vardi (1982) の結果

Vardi (1982) は観測が固定された区間 (すなわち「窓」) からなる場合の, 再生過程について尤度関数を構成し, 寿命分布のノンパラメトリックな最尤推定量を導いた. Vardi (1982) は事象の生起する間隔を離散時間としている. イベントは $a_i + 1, \dots, a_i + b_i$ のいずれかの時点で生起する.

$N_i(j)$ を $a_i + 1, \dots, a_i + j$ ($j = 1, 2, \dots$) の期間で生起したイベントの総数とする. $\xi_i(j)$ を i 番目の過程で a_i 以降に j 番目の失敗が発生した日とする.

Vardi (1982) によれば, 起こり得る観測は次の 4 種類に分類される.

1. $\mathbf{X} = \{\xi_i(j) - \xi_i(j-1); j = 2, \dots, N_i(b_i), N_i(b_i) \geq 2\}$,
2. $\mathbf{Y} = \{\xi_i(1) - a_i(j-1); N_i(b_i) \geq 1\}$,
3. $\mathbf{Z} = \{a_i + b_i + 1 - \xi(N_i(b_i)) \geq 1\}$,
4. $\mathbf{W} = \{b_i + 1; N_i(b_i) = 0\}$.

$t_1 < t_2 < \dots < t_n$ を全体の集合 $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z} \cup \mathbf{W}$ の観測値の取る値とする. また, x_i, y_i, z_i, w_i を $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$ から得られたそれぞれの観測の重複の回数とする.

事象の生起する間隔が有限の平均 μ を持つ分布 $F(x)$ に従うとすると, この枠組みの下で尤度関数は, 以下ようになる.

$$\prod_{i=1}^h \{f(t_i)\}^{x_i} \left\{ \frac{1 - F(t_i)}{\mu} \right\}^{y_i} \{1 - F(t_i)\}^{z_i} \left\{ \sum_{j=t_i}^{\infty} \frac{1 - F(t_i)}{\mu} \right\}. \quad (3.1)$$

3.4.3 Zhao and Nagaraja (2011) の結果

Zhao and Nagaraja (2011) は連続時間の再生過程についての尤度関数を導出し, 再生間隔が指数分布, ワイブル分布, ガンマ分布に従うフィッシャー情報量を求めている. Zhao and Nagaraja (2011) も Vardi (1982) と同様, 観測される再生間隔を 4 種類に分類して尤度を構成している.

N を窓内で観測されるイベントの総数とする. X を観測されるイベントの生起の間隔とする. X は有限の平均 μ を持つ分布 $F(x)$ に従う. Y を窓内で観測される最初のイベントの, 観測の原点からの生起の間隔とする. 2.5 節で述べたように, Y は密度関数,

$$g(y) = \frac{1 - F(y)}{\mu}$$

を持つ分布に従う. Y の生存関数を $G(y)$ で表すことにする. また Z を最後のイベントの生起から, 観測終了時点までの間隔とする. 観測 X, Y, Z を示したものが図 3.3 である.

m を観測が行われる窓の総数とする. n_w を唯一度のイベントの生起も観測されなかった窓の数とする. $m - n_w$ がイベントの生起した窓の個数である. X, Y, Z に観測を区別するための添字 i を付

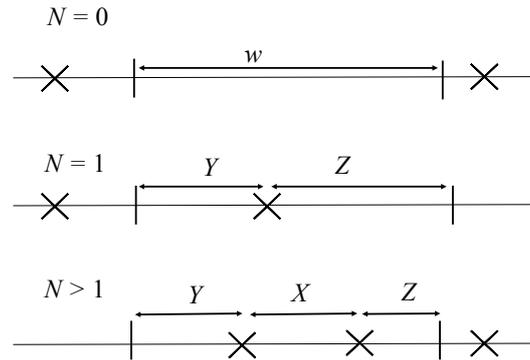


図 3.3 窓打切りされた再生過程の模式図

け, 次の尤度関数が得られる.

$$[G(w)]^{n_w} \prod_{i=1}^{m-n_w} g(y_i) \prod_{i=1}^{n_x} f(x_i) \prod_{i=1}^{m-n_w} [1 - F(z_i)]. \quad (3.2)$$

3.5 マルコフ再生過程を用いたモデル

Zhu *et al.*(2014) は, 建築施設の給水管の故障イベントをモデル化した. 給水管は故障時に修理 (repair) または交換 (replace) を行う. 故障イベントの待ち時間の説明変数として過去の故障回数を用いている. 時刻 T_0 に製品を交換したとし, 時刻 T_1, T_2, \dots, T_k に故障し, 修理を行ったとする. $Y_1 = T_1 - T_0, Y_2 = T_2 - T_1, \dots, Y_k = T_k - T_{k-1}$ は, 正の連続型の分布関数 F_1, F_2, \dots, F_k に従い, 密度関数 f_1, f_2, \dots, f_k を持つ. 交換が行われた場合, 過去の故障回数はリセットされ, 0 に戻る. これはマルコフ再生過程である.

注意する点は, 過去の修理の回数も制限された窓の範囲でしか観測されないという点である.

Zhu *et al.*(2014) ではこの過程を長さ w の窓を通じて観測する. 窓内で観測された故障の回数を $N(w)$ と表記する.

Zhu *et al.*(2014) は K を交換が行われるまでに要した故障の回数を表す確率変数とし確率関数に以下を仮定した.

$$\Pr\{K = k\} = \frac{a^{k-1}}{\sum_{j=0}^{m-1} a^j}, \quad a > 0, \quad k = 1, \dots, m. \quad (3.3)$$

m は k の上界であり, 上界はメンテナンスの方針として予め定まっているものとする. また確率関数 (3.3) のハザードを r_k と表す. すなわち, $r_k = \Pr\{K = k | K \geq k\}$ と置く.

q_k をマルコフ連鎖の定常分布とすると, 以下のように表せる.

$$q_k = \frac{\Pr\{K = k\}}{E[K]} = \frac{\sum_{j=k}^m p_j}{\sum_{j=1}^m j p_j} \quad (3.4)$$

また I を窓の開始点を含む更新間隔の番号を表す確率変数とすると、式 (2.38) の関係より、マルコフ再生過程の定常分布として以下を得る。

$$\Pr(I = k) = \frac{q_k \mu_k}{\sum_{j=1}^m q_j \mu_j}. \quad (3.5)$$

さらに S を窓の開始点を含む更新間隔の長さを表す確率変数とすると、 S の密度関数 ν は以下で与えられる。

$$\nu(s|I = k) = \frac{s f_k(s)}{\mu_k}.$$

ここで μ_k は F_k の平均である。この性質については Cox & Miller (1977) p.356 を参照されたい。 $S = s, I = k$ が与えられたとき、窓の開始点は、区間 $[0, s]$ に一様に分布するため、 t_1 を窓内の最初の故障とすると、

$$g(t_1|S = s, I = k) = \frac{1}{s}, \quad t_1 \in [0, s], \text{ for all } k \quad (3.6)$$

従い、

$$\begin{aligned} g(t_1|I = k) &= \int_{t_1}^{\infty} g(t_1|S = s, I = k) \nu(s|I = k) ds \\ &= \frac{1 - F_k(t_1)}{\mu_k}. \end{aligned} \quad (3.7)$$

尤度関数は複雑な形になるため、いくつかの例を示そう。

まず観測期間内にただ1度も故障が起きなかった場合、(3.7) 式および (3.5) 式より以下の式が成り立つ。

$$\begin{aligned} \Pr\{N(w) = 0\} &= \sum_{k=1}^m \Pr\{T_1 > w|I = k\} \Pr\{I = k\} \\ &= \sum_{k=1}^m \int_w^{\infty} \frac{1 - F_k(t_1)}{\mu_k} dt_1 \cdot \frac{q_k \mu_k}{\sum_{j=1}^m q_j \mu_j} \\ &= \sum_{k=1}^m \frac{q_k}{\sum_{j=1}^m q_j \mu_j} \int_w^{\infty} \{1 - F_k(t_1)\} dt_1. \end{aligned} \quad (3.8)$$

次に、観測期間内に1度だけ故障が起きた場合の尤度を考える。故障は区間 $[t_1, t_1 + \delta t]$ に生じたとする。 \mathfrak{R}_i を t_i で交換がおこなわれた場合に1、さもなければ0の値を取る変数とする。

修理が1度だけ行われた場合の尤度は、

$$\begin{aligned} \Pr\{T_1 \in [t_1, t_1 + \delta t] \cap \mathfrak{R}_1 = 0 \cap T_2 > w\} &= \sum_{k=1}^{m-1} \Pr\{T_1 \in [t_1, t_1 + \delta t]|I = k\} \Pr\{I = k\} \\ &= \sum_{k=1}^{m-1} \frac{q_k}{\sum_{j=1}^m q_j \mu_j} \{1 - F_k(t_1)\} \times (1 - r_k) \{1 - F_{k+1}(w - t_1)\} \delta t \end{aligned}$$

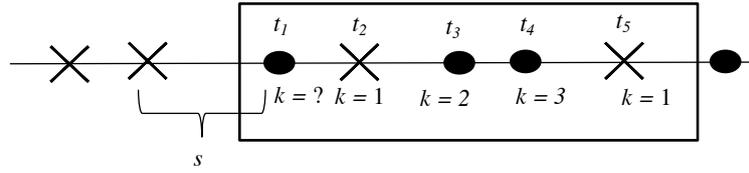


図 3.4 観測される故障のパターンの例. ●は修理, ×は交換を表す.

交換が1度だけ行われた場合の尤度は,

$$\Pr\{T_1 \in [t_1, t_1 + \delta t] \cap \mathfrak{R}_0 = 1 \cap T_2 > w\} = \sum_{k=1}^m \frac{q_k}{\sum_{j=1}^m q_j \mu_j} \{1 - F_k(t_1)\} r_k \{1 - F_1(w - t_1)\} \delta t.$$

同様に、複数回の故障が起こった場合の尤度も構成できる. 図 3.4 のように、{修理, 交換, 修理, 修理, 交換} が t_1, \dots, t_5 の時点で生じた場合、

$$\begin{aligned} & \sum_{k=1}^{m-1} \left(\frac{q_k}{\sum_{j=1}^m q_j \mu_j} \{1 - F_k(t_1)\} (1 - r_k) f_{k+1}(t_2 - t_1) r_{k+1} \right) \\ & \times f_1(t_3 - t_2) (1 - r_1) f_2(t_4 - t_3) (1 - r_3) f_3(t_5 - t_4) r_3 \\ & \times \{1 - F_1(w - t_5)\} \delta t_1 \delta t_2 \delta t_3 \delta t_4 \delta t_5. \end{aligned}$$

さて、尤度関数をまとめると以下の様になる. $t_{j_{\mathfrak{R}}}$ を観測範囲内での最初の交換の起こった時点として、

$$L(\theta; d, w) = \begin{cases} \sum_{i=1}^{m-n} \frac{q_i}{\sum_{j=i}^m q_j \mu_j} \prod_{j=1}^{n+1} L_{j, i+j-1} & \text{交換のない場合,} \\ \sum_{i=1}^{m-j_{\mathfrak{R}}+1} \frac{q_i}{\sum_{j=i}^m q_j \mu_j} \prod_{j=1}^{j_{\mathfrak{R}}} L_{j, i+j-1} \prod_{j=j_{\mathfrak{R}}+1}^{n+1} L_{j, k_j} & \text{交換が少なくとも1回ある場合,} \end{cases}$$

ここで $L_{j,k}$ は次のように与えられる.

$$L_{1,k} = \begin{cases} \{1 - F_k(t_1)\} (1 - r_k), & \text{最初の故障で修理が行われた場合,} \\ \{1 - F_k(t_1)\} r_k, & \text{最初の故障で交換が行われた場合,} \end{cases}$$

$$L_{j,k} = \begin{cases} f_k(t_j - t_{j-1}) (1 - r_k) & j \text{ 番目 } (1 < j \leq n) \text{ の故障で修理が行われた場合,} \\ f_k(t_j - t_{j-1}) r_k & j \text{ 番目 } (1 < j \leq n) \text{ の故障で交換が行われた場合,} \end{cases}$$

$$L_{n+1,k} = 1 - F_k(w - t).$$

ただし観測期間内にただ一度の故障も起きなかったとき、尤度は (3.8) 式で与えられる。

さらに独立な M 個のサンプルパスが同じ過程から観測されたとする。

データ・セット $D = d_1, \dots, d_M$ の結合尤度関数は、

$$L(\theta; D) = \prod_{s=1}^M L(\theta; d_s, w_s). \quad (3.9)$$

(3.9) 式を最大化することでパラメータを求める。

3.6 窓打切りされた線分過程のパラメータ推定

3.6.1 線分過程

我々の目的は顕微鏡写真からヒートシール界面の長さの分布のパラメータの推定を行うことである。ヒートシール界面は線分とみなすことができる。そのためこの節では、1次元の空間上にランダムに配置された線分について考察する。ランダムな線分の開始点 S と線分の長さ L の組の変数 $\{S_i, L_i\} (i = 1, 2, \dots)$ の集合を線分過程 (line segment process) と呼ぶ。線分過程は Parker & Cowan (1976), Laslett (1982) で研究された。Parker & Cowan(1976) は線分の中点がポアソン過程に従って配置された場合の、線分の長さや数の期待値を導出した。Laslett (1982) は線分の開始点が定常ポアソン過程に従う場合の、線分の長さの分布の推定量を提案した。Laslett (1982) はまた、生存時間分析と空間上の線分の長さの分布を推定する問題の共通性を指摘している。Mauldon(1998) は定常ポアソンの仮定をおかず、より一般的な枠組のもとで線分の長さや数の期待値の推定量を提案した。我々は Laslett(1982) および Mauldon(1998) の推定量について、シミュレーションによる比較を行う。

3.6.2 Laslett (1982) の結果

生存時間分析の分野では、観測されたデータに打切りがある場合の最尤推定について研究がなされている。まず観察期間 $(T_0, T_0 + w)$ 内のランダムな時間に到着した m 人の患者の生存時間分布の推定を考える。観測されるのは変数の組 $(Z_i = z_i, D_i = d_i)$, $(i = 1, \dots, m)$ である。 Z_i はそれぞれ独立に同分布 F に従うとする。 D_i は右打切りの有無を表し、観測が打ち切られた場合に 0、そうでなければ 1 の値をとる。打切りの発生は Z_i と独立に生じるものとする。 $d_i = 0$ のとき、観測値 z_i が示すのは、患者 i の生存時間が z_i 以上であるということである。よってこの場合、尤度は (3.2) 節で述べた通り、

$$\prod_{i=1}^m \{f(z_i)\}^{d_i} \{1 - F(z_i)\}^{1-d_i} \quad (3.10)$$

となる。

次に、観察の原点 T_0 の前に到着した l 人の患者の、 T_0 からの生存時間も観測されているものとする。これを $(Y_j = y_j, E_j = e_j)$, $(i = 1, \dots, l)$ と表す。 E_i は右打切りの有無を表し、観測が打ち切ら

れた場合に 0, そうでなければ 1 の値をとる. 標本の大きさは $N = m + l$ である. この場合, 時刻 T_0 で生存している患者の余命が y 以上の割合は, o が十分大きいとき, 再生過程における均衡分布と同様,

$$G(y) = \int_y^\infty \frac{1 - F(u)}{\mu} du \quad (3.11)$$

となる. ここで μ は分布 F の平均を表す. G は生存関数であるから, 密度関数は $g(y) = -\frac{d}{dw}G(y) = (1 - F(y))/\mu$ である.

Laslett (1982) では, 患者の到着が定常ポアソン過程に従うことを仮定した. このとき l がパラメータ $N = m + l$, $\mu/(\mu + w)$ の二項分布に従うことを指摘した上で, 上記の状況で最大化すべき尤度はこれらをすべて掛け合わせたものになると述べている. すなわち尤度 L は,

$$L = \binom{m+l}{l} \left(\frac{\mu}{\mu+w}\right)^l \left(\frac{w}{\mu+w}\right)^m \cdot \prod_{i=1}^m \{f(z_i)\}^{d_i} \{1 - F(z_i)\}^{1-d_i} \prod_{j=1}^l \{g(y_j)\}^{e_j} \{G(y_j)\}^{1-e_j}, \quad (3.12)$$

である.

数値計算上都合のよいように, (3.12) 式の記号を $(\tau_1, \dots, \tau_N) = (z_1, \dots, z_m, y_1, \dots, y_l)$, $(\delta_1^{(r)}, \dots, \delta_N^{(r)}) = (e_1, \dots, e_l, d_1, \dots, d_m)$ と付け変える. また $\delta_i^{(l)}$ ($i = 1, \dots, N$) なる変数を導入し, 患者 k が観測開始時刻 T_0 の前に到着した場合に 0, そうでなければ 1 の値をとるものとする.

$$L = \prod_{i=1}^N \left(\frac{\mu}{\mu+w}\right)^{1-\delta_i^{(l)}} \left(\frac{w}{\mu+w}\right)^{\delta_i^{(l)}} \left[\{f(\tau_i)\}^{\delta_i^{(r)}} \{1 - F(\tau_i)\}^{1-\delta_i^{(r)}} \right]^{\delta_i^{(l)}} \left[\{g(\tau_i)\}^{\delta_i^{(r)}} \{G(\tau_i)\}^{1-\delta_i^{(r)}} \right]^{1-\delta_i^{(l)}}. \quad (3.13)$$

μ の推定量は (3.13) 式を最大化することにより求まる.

また N はパラメータ $\lambda(\mu + w)$ のポアソン分布に従うため, λ の最尤推定量は $N/(\mu + w)$ となる. μ の最尤推定量を代入し, 以下の推定量を用いることができる.

$$\hat{\lambda} = \frac{N}{\mu + w} \quad (3.14)$$

Laslett(1982) では, 尤度関数の導出の細部は省略されているため, ここで改めて整理する. まず式 (3.11) が以下にして導かれるか示そう. 線分の長さに対応する確率変数を X , その分布を $F(x)$, $F(x)$ の平均を μ 確率変数 X の実現値を x とする. 区間 $(0, T_0)$ 内の点 U に到着した患者を考える. 定常ポアソンの仮定より U は一様分布する. $U = u$ が与えられたとき, この患者の生存時間が観測の開始点 T_0 に到達する確率は,

$$P(X + u > T_0) = 1 - F(T_0 - u) \quad (3.15)$$

周辺化し,

$$p_{T_0} = \frac{1}{T_0} \int_0^{T_0} \{1 - F(u)\} du \quad (3.16)$$

区間点 T_0 に n 人の患者が到着し、そのうち i 人の生存時間が T_0 に到達する確率は、

$$\sum_{n=i}^{\infty} \frac{(\lambda T_0)^n e^{-\lambda T_0}}{n!} \binom{n}{i} p_{T_0}^i (1 - P_{T_0})^{n-i}. \quad (3.17)$$

であり、

$$\left(\lambda \int_0^{T_0} \{1 - F(u)\} du \right)^i e^{-\lambda \int_0^{T_0} \{1 - F(u)\} du} \quad (3.18)$$

T_0 を無限大とすると、平均 $\lambda\mu$ のポアソン分布になる。原点をまたぐ生存時間の確率密度は、線分の長さにも比例するため、原点をまたぐ生存時間の X の確率密度は、

$$xf(x)/\mu. \quad (3.19)$$

原点は長さ x の区間内に一様に分布する。原点から観測された生存時間の残りの長さ Y と生存時間の長さ X の同時密度は、

$$\frac{1}{x} xf(x)/\mu, \quad (3.20)$$

X について積分し、以下の結果を得る。

$$g(x) = \int_y^{\infty} \frac{f(x)}{\mu} dx = \frac{1 - F(y)}{\mu}. \quad (3.21)$$

この事実に関しては Feller (1971), Smith (1958) でも論じられている。

次に l, m, N それぞれの分布と、 N を所与としたときの l の分布を導く。

l はパラメータ $\mathbb{E}[m] = \lambda\mu$ のポアソン分布に従うことを示す。区間 $(0, T_0)$ 上の点 U に到着した患者の生存時間が、 T_0 まで続いている確率は、以下の式で表せる。

$$p(t) = \frac{1}{t} \int_0^t \{1 - F(t - u)\} du.$$

区間 $(0, t)$ 内に n 人の患者が到着したとすると、このうち i 人の生存時間が T_0 まで続く確率は、

$$\binom{n}{i} p(t)^i (1 - p(t))^{n-i}.$$

点 T_0 をまたぐ生存時間の数が i である確率は、

$$\sum_{n=i}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \binom{n}{i} p(t)^i (1 - p(t))^{n-i}$$

整理して $t \rightarrow \infty$ とすると、

$$\frac{(\lambda\mu)^i}{i!} e^{-\lambda\mu}. \quad (3.22)$$

これはポアソン分布の確率関数である。

m については、2.3 節での議論からすぐに、パラメータ $\mathbb{E}[l] = \lambda w$ のポアソン分布従うことがわかる。

それぞれパラメータ $\lambda\mu, \lambda w$ を 2 つのポアソン分布の和の分布はパラメータ $\lambda(\mu + w)$ を持つポアソン分布となる。 $N = m + l$ が与えられたとき、 l はパラメータ $\mu/(\mu + w)$ を持つ二項分布に従う。条件付き確率の定義より、

$$\begin{aligned} P(l = k | N = j) &= \left(\frac{(\lambda\mu)^k \exp(-\lambda\mu)}{k!} \right) \left(\frac{(\lambda w)^{j-k} \exp(-\lambda w)}{(j-k)!} \right) / \left(\frac{(\lambda(\mu + w))^j \exp(-\lambda(\mu + w))}{j!} \right) \\ &= \binom{j}{k} \left(\frac{\lambda}{\lambda + w} \right)^k \left(\frac{w}{\lambda + w} \right)^{j-k}. \end{aligned}$$

となりこれは二項分布の確率関数である。

以上の議論により、(3.13) 式に示した尤度関数が導かれる。

指数分布の場合

Laslett (1982) はノンパラメトリックな生存関数の推定に関心があるため、具体的な分布を仮定しての議論は行っていない。寿命分布が指数分布に従うについては、推定量が閉じた形で求まる。指数分布の場合、均衡分布の密度関数と元の密度関数が一致するという性質があるため、尤度関数は以下のようになる。

$$L = \binom{N}{n_1} \left(\frac{\mu}{\mu + w} \right)^{n_1} \left(\frac{w}{\mu + w} \right)^{N-n_1} \prod_{i=1}^N \{f(z_i)\}^{\delta_i^{(r)}} \{1 - F(z_i)\}^{1-\delta_i^{(r)}} \quad (3.23)$$

$$\propto \left(\frac{\mu}{\mu + w} \right)^{n_1} \left(\frac{w}{\mu + w} \right)^{N-n_1} \prod_{i=1}^N \left\{ \frac{1}{\mu} \exp\left(-\frac{z_i}{\mu}\right) \right\}^{\delta_i^{(r)}} \left\{ \exp\left(-\frac{z_i}{\mu}\right) \right\}^{1-\delta_i^{(r)}} \quad (3.24)$$

対数を取り微分すると以下の式を得る。

$$\frac{d}{d\mu} \log L = \frac{n_1}{\mu} - \frac{N}{\mu + w} + \sum_{i=1}^N \left\{ \frac{z_i}{\mu^2} - \frac{d_i}{\mu} \right\}. \quad (3.25)$$

これを 0 と置くと μ についての 2 次方程式となるため、

$$\begin{aligned} A &= n_1 - N - \sum_{i=1}^N \delta_i^{(r)}, \\ B &= \sum_{i=1}^N y_i - w \sum_{i=1}^N \delta_i^{(r)} + n_1 w, \\ C &= w \sum_{i=1}^N y_i, \end{aligned}$$

と置くと、解は

$$\hat{\mu} = \frac{-B - \sqrt{B^2 - 4AC}}{2A} \quad (3.26)$$

である。すべての観測が両側打切りのときは解を持たない。

Laslett (1982) の提案した推定量は、最尤推定の枠組みに基づくものであるため、適当な分布を仮定すればフィッシャー情報行列が漸近分散を導くことができる。対数尤度関数の2階微分を求めると以下ようになる。

$$\frac{d^2}{d\mu^2} \log L = -\frac{n_1}{\mu^2} + \frac{N}{(\mu+w)^2} + \sum_{i=1}^N \left\{ -2\frac{z_i}{\mu^3} + \frac{\delta_i^{(r)}}{\mu^2} \right\}. \quad (3.27)$$

負の期待値を取り、フィッシャー情報量を求める。

$$-E \left[\frac{d^2}{d\mu^2} \log L \right] = \frac{\lambda(2\mu w + w^2)}{\mu^2(\mu + w)}. \quad (3.28)$$

以上の議論により、以下の事実が導かれる。

$$\hat{\mu} \xrightarrow{d} \text{Normal} \left(\mu, \frac{\mu^2(\mu+w)}{\lambda(2\mu w + w^2)} \right). \quad (3.29)$$

ここで期待値の計算には以下の性質を用いた。

- $E[n_1] = \lambda\mu$ (3.22) 式を参照)
- $E[N - n_1] = \lambda w$ (2.3 節を参照)
- $E[\sum \delta_i^{(r)}] = \lambda w$ (2.3 節を参照)
- $E[\sum z_i] = \lambda\mu w$ (Parker and Cowan, 1976 を参照)

これらの期待値の導出については、Parker and Cowan (1976) が詳しい。

信頼区間の構成

生存時間が指数分布に従う場合、式 (3.29) の関係から正規分布を用いた近似信頼区間を導くことができる。ワイブル分布やガンマ分布に従う場合は、対数尤度関数を数値的に最大化する際に副産物として得られる近似ヘシアン (観測情報行列) から推定値の標準誤差を求め、それを用いて信頼区間を構成する。

3.6.3 Mauldon (1998) の推定量

Laslett (1982) は生存時間の分布を推定することに関心を置いていた。Mauldon (1998) は、岩石のトレースの長さや数の平均の推定量を提案した。

まず Mauldon (1998) が扱った問題を整理する。Mauldon (1998) は岩石のトレースが2次元空間上の平行な線分として近似できるとしている。線分は矩形の窓を通して観察される。線分どうしが平行であるため、 y 軸の情報は無視できる。矩形の窓の幅を w と置き、高さを 1 と置いて一般性を失わない (図 3.5)。

Laslett (1988) は、生存時間を扱う問題と線分の長さを扱う問題の同一性を指摘している。3.6.2 節で生存時間として扱った対象を、空間上の線分と捉えることが可能である。ただし Laslett (1982) と異なり Mauldon (1998) は線分の始点にポアソン過程を仮定せず、一般の再生過程を仮定している。

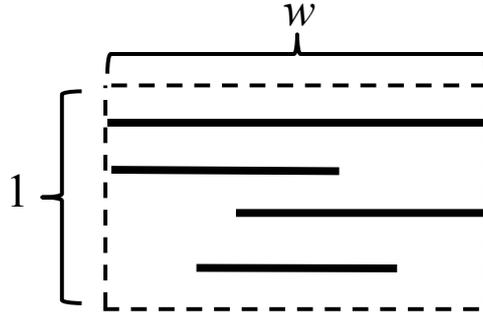


図 3.5 Mauldon(1998) が扱った観測の模式図

λ の推定量

λ を単位区間あたりの線分の中点の数の平均とする。中点の数と始点の数は一致するため、中点の数の平均を推定する問題と始点の数の平均を推定する問題は同じである。z_i をトレースの長さとし、F(z) をトレースの長さの分布とする。F(z) は有限の平均 μ を持つ。N を窓を通して少なくとも部分的に観測されるトレースの総数とする。N_T を両側で打ち切られたトレースの数とする。N_C を打ち切られていないトレースの数とする。このとき、観測される線分の端点の数は（始点と終点の 2 つがあるため）中点の数の 2 倍に等しい。ここで窓の中で観測される端点の数は、(N - N_T - N_C) + 2N_C = (N - N_T + N_C) である。これより推定量、

$$\tilde{\lambda} = \frac{1}{2w} (N - N_T + N_C) \tag{3.30}$$

が得られる。

3.6.4 μ の推定量 (1)

次に μ の推定量を考える。Mauldon(1998) は μ に対する 2 種類の推定量を提案した。一般に E[N] = E[E[N|z]] であるから、

$$E[N] = \int_0^\infty E[N|z] dF(z). \tag{3.31}$$

いま、z を固定して考えると、E[N|t] = λ(z + w) であるから、

$$E[N] = \int_0^\infty \lambda(z + w) dF(z) = \lambda \int_0^\infty z dF(z) + \lambda w \int_0^\infty dF(z) \tag{3.32}$$

$$= \lambda(\mu + w). \tag{3.33}$$

これより、

$$\mu = E[N]/\lambda - w \tag{3.34}$$

を得る. (3.34) 式において, $E[N]$ を N で, λ を $\tilde{\lambda}$ で置き換えることにより, 次の推定量を得る.

$$\tilde{\mu}^{(1)} = w \left(\frac{N + N_T - N_C}{N - N_T + N_C} \right). \quad (3.35)$$

3.6.5 μ の推定量 (2)

$N(w)$ を長さ w の窓を通して観測されるトレースの総数とする. これまでの議論より, $E[N(0)] = \lambda\mu$ であり, $E[N(w)] = \lambda(\mu + w)$ である.

$$\frac{E[N(0)]}{\mu} = \frac{E[N(w)]}{\mu + w}. \quad (3.36)$$

$$\frac{1}{\mu} = \frac{1}{h} \frac{E[N(h)] - E[N(0)]}{E[N(0)]}. \quad (3.37)$$

\bar{y} を観測範囲内のトレースの長さの平均とする. 観測されるトレースの長さの総和を L とする. いま,

$$N(0) = \frac{L}{h} = \frac{\bar{y}N(h)}{h} \quad (3.38)$$

これより推定量,

$$\frac{1}{\tilde{\mu}^{(2)}} = \frac{1}{\bar{y}} - \frac{1}{h} \quad (3.39)$$

を得る.

3.6.6 推定量の比較

線分の始点がポアソン過程に従うとした場合, Mauldon (1998) の推定量と Laslett (1982) の推定量は比較が可能であるが, そのような考察はこれまでなされていない. 本節ではシミュレーションにより, Mauldon (1998) の推定量と Laslett (1982) の推定量の性質を調べる. はじめに Mauldon (1998) および Laslett (1982) の仮定した状況をシミュレーションで再現するアルゴリズム (1) に示す. このシミュレーションでは, 広い区間 $[0, 1000]$ に発生させた線分のうち狭い区間 $[500, 501]$ に入ったもののみを切り出すことを行う.

窓の幅を固定し, すべての線分が平行であると仮定しているため, 観測範囲内での線分の始点 S_i と終点 E_i が得られれば, Mauldon (1998) および Laslett (1982) の推定量を利用するための情報を失わない. すなわち, 観測 (S_i, E_i) の組から以下のようにインジケータ変数 $\delta_i^{(l)}, \delta_i^{(r)}$ が以下のように得られる.

- 打ち切りなしの場合: $S_i > 0, E_i < w$ のとき, $\delta_i^{(l)} = 1, \delta_i^{(r)} = 1$
- 右打ち切りの場合: $S_i > 0, E_i = w$ のとき, $\delta_i^{(l)} = 1, \delta_i^{(r)} = 0$
- 左打ち切りの場合: $S_i = 0, E_i < w$ のとき, $\delta_i^{(l)} = 0, \delta_i^{(r)} = 1$
- 両側切りの場合: $S_i = 0, E_i = w$ のとき, $\delta_i^{(l)} = 0, \delta_i^{(r)} = 0$

Algorithm 1 線分過程をシミュレートするアルゴリズム**Require:** ポアソン過程および $f_1(x)$, $f_0(x)$ のパラメータ.**Ensure:** S_j, E_j , ($i = 1, \dots, m$). $s_i \sim \text{Uniform}(0, 1000)$, $r_i \sim \text{Exponential}(\mu)$ $e_i = p_i + r_i$ **if** $j \in \{j | (500 < e_j < 501) \cup (500 < s_j < 501) \cup (s_j < 500 \cap 501 < e_j)\}$ **then** $S_j = \max(s_j - 500, 0)$, $E_j = \min(e_j - 500, 1)$ **end if**

Laslett(1982)の方法について指数分布の最尤推定量は式(3.26)の示した通り、閉じた形で求まる。ワイブル分布、ガンマ分布を仮定した場合、尤度関数を数値的に最大化することにより推定を行う。今回はネルダー・ミード法(Nelder & Mead, 1965)を用いた。

Mauldon(1998)の方法では、分布にかかわらず閉じた形で推定量が得られている。

10000回のシミュレーションを行った結果、推定値の平均を表3.1-3.3に、RMSEを表3.4-3.6に示す。正規分布を用いた近似信頼区間の被覆確率(coverage probability)については表3.7-3.9に示す。

表3.1 推定値の平均の比較. 指数分布を仮定した場合.

λ	μ	$\hat{\lambda}$	$\tilde{\lambda}$	$\hat{\mu}$	$\mu^{(1)}$	$\mu^{(2)}$
50	0.10	50.08	50.06	0.10	0.10	0.10
50	0.50	49.98	49.97	0.50	0.50	0.50
50	1.00	50.02	50.03	1.01	1.01	1.01
50	2.00	49.99	49.99	2.02	2.02	2.02
100	0.10	100.00	100.01	0.10	0.10	0.10
100	0.50	99.90	99.90	0.50	0.50	0.50
100	1.00	99.95	99.96	1.00	1.00	1.01
100	2.00	100.07	100.06	2.01	2.01	2.01
200	0.10	199.79	199.74	0.10	0.10	0.10
200	0.50	200.02	200.00	0.50	0.50	0.50
200	1.00	200.13	200.14	1.00	1.00	1.00
200	2.00	199.93	199.93	2.00	2.00	2.01

バイアスに関してはいずれの方法を用いても0に近いことがわかる。RMSEに関してはLaslett(1982)の方法が若干小さい値を取る。これはLaslett(1982)の方法では分布に関する情報を使えるためであろう。被覆確率については、指数分布を仮定した場合、名目上の値0.95に近い値を取る。ワイブル分布、ガンマ分布の場合には、形状母数については名目上の値0.95に近い値を取るが、尺度母数については λ が小さい場合、被覆確率が名目上の値よりもやや小さくなる。本節では平均

表 3.2 推定値の平均の比較. ワイブル分布を仮定した場合.

m	λ	μ	$\hat{\lambda}$	$\tilde{\lambda}$	$\hat{\mu}$	$\tilde{\mu}^{(1)}$	$\tilde{\mu}^{(2)}$
0.50	50.00	0.10	49.94	49.94	0.10	0.10	0.10
0.50	50.00	0.50	50.06	50.05	0.50	0.51	0.51
0.50	50.00	1.00	50.12	50.11	1.01	1.01	1.01
0.50	50.00	2.00	50.05	50.07	2.02	2.02	2.02
0.50	100.00	0.10	100.06	100.06	0.10	0.10	0.10
0.50	100.00	0.50	99.99	99.97	0.50	0.50	0.50
0.50	100.00	1.00	100.13	100.12	1.00	1.01	1.01
0.50	100.00	2.00	100.03	100.01	2.01	2.01	2.01
0.50	200.00	0.10	200.18	200.18	0.10	0.10	0.10
0.50	200.00	0.50	200.07	200.09	0.50	0.50	0.50
0.50	200.00	1.00	200.21	200.24	1.00	1.00	1.00
0.50	200.00	2.00	200.06	200.08	2.01	2.01	2.01
2.00	50.00	0.10	50.03	50.04	0.10	0.10	0.10
2.00	50.00	0.50	49.88	49.87	0.50	0.50	0.50
2.00	50.00	1.00	49.99	49.99	1.00	1.00	1.01
2.00	50.00	2.00	49.98	49.96	2.01	2.01	2.02
2.00	100.00	0.10	99.98	100.00	0.10	0.10	0.10
2.00	100.00	0.50	99.98	99.99	0.50	0.50	0.50
2.00	100.00	1.00	100.05	100.07	1.00	1.00	1.01
2.00	100.00	2.00	99.96	99.96	2.01	2.01	2.01
2.00	200.00	0.10	199.94	199.94	0.10	0.10	0.10
2.00	200.00	0.50	199.94	199.91	0.50	0.50	0.50
2.00	200.00	1.00	199.88	199.88	1.00	1.00	1.00
2.00	200.00	2.00	199.94	199.93	2.00	2.00	2.00

の推定結果のみを記載した. 各パラメータの推定値の平均と標準誤差は付録に記載する.

3.7 この章のまとめ

この章では窓打切りされた観測からパラメータを推定する方法をまとめ, まだ十分に推定量の性質が明らかになっていない場合についてはシミュレーションを用いて検証を行った. 最尤推定を行う場合には, 観測されたイベントの生起の間隔を, 完全な観測, 窓の終点のみで途切れた観測, 窓の始点のみで途切れた観測, 窓の終点と始点, 両側で途切れた観測の4種類に分類するアプローチが有用であることがわかった.

加えて Laslett(1982) の議論から, 尤度関数を正しく導くためには打ち切られた観測の個数も考慮

表 3.3 推定値の平均の比較. ガンマ分布を仮定した場合.

k	λ	μ	$\hat{\lambda}$	$\tilde{\lambda}$	$\hat{\mu}$	$\tilde{\mu}^{(1)}$	$\tilde{\mu}^{(2)}$
0.50	50.00	0.10	50.02	50.02	0.10	0.10	0.10
0.50	50.00	0.50	50.04	50.04	0.50	0.51	0.51
0.50	50.00	1.00	49.96	49.98	1.01	1.01	1.01
0.50	50.00	2.00	50.01	50.04	2.02	2.02	2.02
0.50	100.00	0.10	99.91	99.91	0.10	0.10	0.10
0.50	100.00	0.50	99.94	99.96	0.50	0.50	0.50
0.50	100.00	1.00	100.06	100.03	1.01	1.01	1.00
0.50	100.00	2.00	100.05	100.09	2.01	2.01	2.01
0.50	200.00	0.10	199.70	199.74	0.10	0.10	0.10
0.50	200.00	0.50	199.94	199.94	0.50	0.50	0.50
0.50	200.00	1.00	200.01	200.03	1.00	1.00	1.00
0.50	200.00	2.00	200.09	200.09	2.00	2.00	2.00
2.00	50.00	0.10	49.92	49.92	0.10	0.10	0.10
2.00	50.00	0.50	49.96	49.95	0.50	0.50	0.50
2.00	50.00	1.00	49.99	49.98	1.01	1.01	1.01
2.00	50.00	2.00	49.85	49.83	2.01	2.02	2.02
2.00	100.00	0.10	99.95	99.97	0.10	0.10	0.10
2.00	100.00	0.50	100.15	100.15	0.50	0.50	0.50
2.00	100.00	1.00	100.15	100.16	1.00	1.00	1.00
2.00	100.00	2.00	99.96	99.94	2.01	2.01	2.01
2.00	200.00	0.10	199.89	199.87	0.10	0.10	0.10
2.00	200.00	0.50	200.02	200.01	0.50	0.50	0.50
2.00	200.00	1.00	199.76	199.73	1.00	1.00	1.00
2.00	200.00	2.00	200.16	200.15	2.00	2.00	2.01

しなければならないことがわかる.

線分過程のパラメータ推定に関しては Laslett(1982) の方法と Mauldon(1998) の推定量の比較を行い, 分布が正しく仮定できる場合には RMSE の面で Laslett(1982) の方法が優れていることがわかった. しかし, Laslett(1982) の方法は線分の生起の間隔に定常ポアソン過程を仮定しており, 一般の場合に窓打切り状況下で線分過程の分布のパラメータを推定する方法の開発は今後の課題といえるであろう.

また Zhu *et al.*(2014) の方法について論じた 3.5 節で見たように, 窓打切りデータに基づく尤度関数はしばしば非常に複雑な形となる. これに対してはより計算を簡略化するため, EM アルゴリズムの適用も検討する余地がある.

表 3.4 推定値の RMSE の比較. 指数分布を仮定した場合.

λ	μ	$\hat{\lambda}$	$\tilde{\lambda}$	$\hat{\mu}$	$\mu^{(1)}$	$\mu^{(2)}$
50.00	0.10	6.74	6.88	0.01	0.03	0.01
50.00	0.50	6.13	6.26	0.06	0.07	0.07
50.00	1.00	5.77	5.85	0.12	0.12	0.14
50.00	2.00	5.47	5.50	0.22	0.22	0.26
100.00	0.10	9.65	9.80	0.01	0.02	0.01
100.00	0.50	8.67	8.86	0.04	0.05	0.05
100.00	1.00	8.15	8.24	0.08	0.09	0.10
100.00	2.00	7.78	7.82	0.16	0.16	0.19
200.00	0.10	13.59	13.81	0.01	0.02	0.01
200.00	0.50	12.33	12.62	0.03	0.04	0.04
200.00	1.00	11.43	11.59	0.06	0.06	0.07
200.00	2.00	10.77	10.83	0.11	0.11	0.13

表 3.5 推定値の RMSE の比較. ワイブル分布を仮定した場合.

m	λ	μ	$\hat{\lambda}$	$\tilde{\lambda}$	$\hat{\mu}$	$\tilde{\mu}^{(1)}$	$\tilde{\mu}^{(2)}$
0.50	50.00	0.10	6.88	6.96	0.02	0.03	0.03
0.50	50.00	0.50	6.41	6.47	0.09	0.09	0.09
0.50	50.00	1.00	6.17	6.24	0.15	0.16	0.17
0.50	50.00	2.00	5.98	6.04	0.27	0.27	0.29
0.50	100.00	0.10	9.70	9.79	0.02	0.02	0.02
0.50	100.00	0.50	9.06	9.21	0.06	0.06	0.07
0.50	100.00	1.00	8.79	8.87	0.11	0.11	0.12
0.50	100.00	2.00	8.55	8.61	0.19	0.20	0.21
0.50	200.00	0.10	13.73	13.89	0.01	0.02	0.01
0.50	200.00	0.50	12.81	12.96	0.04	0.05	0.05
0.50	200.00	1.00	12.39	12.52	0.07	0.08	0.08
0.50	200.00	2.00	11.94	12.05	0.13	0.14	0.14
2.00	50.00	0.10	6.72	6.88	0.01	0.03	0.01
2.00	50.00	0.50	5.97	6.22	0.04	0.06	0.06
2.00	50.00	1.00	5.40	5.45	0.09	0.09	0.13
2.00	50.00	2.00	5.16	5.17	0.20	0.20	0.26
2.00	100.00	0.10	9.56	9.77	0.01	0.02	0.01
2.00	100.00	0.50	8.47	8.74	0.03	0.04	0.04
2.00	100.00	1.00	7.63	7.69	0.06	0.06	0.09
2.00	100.00	2.00	7.29	7.29	0.13	0.13	0.18
2.00	200.00	0.10	13.52	13.76	0.00	0.02	0.00
2.00	200.00	0.50	11.84	12.28	0.02	0.03	0.03
2.00	200.00	1.00	10.99	11.13	0.04	0.05	0.06
2.00	200.00	2.00	10.36	10.37	0.09	0.10	0.13

表 3.6 推定値の RMSE の比較. ガンマ分布を仮定した場合.

k	λ	μ	$\hat{\lambda}$	$\tilde{\lambda}$	$\hat{\mu}$	$\tilde{\mu}^{(1)}$	$\tilde{\mu}^{(2)}$
0.50	50.00	0.10	6.78	6.88	0.02	0.03	0.02
0.50	50.00	0.50	6.28	6.39	0.08	0.08	0.09
0.50	50.00	1.00	6.12	6.18	0.14	0.15	0.16
0.50	50.00	2.00	5.77	5.81	0.25	0.26	0.28
0.50	100.00	0.10	9.75	9.87	0.01	0.02	0.01
0.50	100.00	0.50	8.99	9.16	0.05	0.06	0.06
0.50	100.00	1.00	8.53	8.61	0.10	0.10	0.11
0.50	100.00	2.00	8.13	8.19	0.18	0.18	0.20
0.50	200.00	0.10	13.48	13.69	0.01	0.02	0.01
0.50	200.00	0.50	12.58	12.81	0.04	0.04	0.04
0.50	200.00	1.00	12.12	12.25	0.07	0.07	0.08
0.50	200.00	2.00	11.70	11.77	0.12	0.13	0.14
2.00	50.00	0.10	6.70	6.84	0.01	0.03	0.01
2.00	50.00	0.50	6.00	6.17	0.05	0.07	0.06
2.00	50.00	1.00	5.57	5.65	0.10	0.11	0.13
2.00	50.00	2.00	5.22	5.25	0.20	0.21	0.26
2.00	100.00	0.10	9.59	9.80	0.01	0.02	0.01
2.00	100.00	0.50	8.46	8.75	0.03	0.05	0.04
2.00	100.00	1.00	7.89	7.97	0.07	0.07	0.09
2.00	100.00	2.00	7.35	7.38	0.14	0.14	0.18
2.00	200.00	0.10	13.71	13.97	0.00	0.02	0.01
2.00	200.00	0.50	12.18	12.50	0.02	0.03	0.03
2.00	200.00	1.00	11.28	11.48	0.05	0.05	0.06
2.00	200.00	2.00	10.52	10.56	0.10	0.10	0.13

表 3.7 μ の被覆確率. 指数分布を仮定した場合

λ	μ	μ の被覆確率
50.00	0.10	0.94
50.00	0.50	0.94
50.00	1.00	0.95
50.00	2.00	0.95
100.00	0.10	0.94
100.00	0.50	0.95
100.00	1.00	0.95
100.00	2.00	0.95
200.00	0.10	0.94
200.00	0.50	0.95
200.00	1.00	0.95
200.00	2.00	0.95

表 3.8 パラメータの被覆確率. ワイブル分布を仮定した場合

λ	m	η	m の被覆確率	η の被覆確率
50.00	0.50	0.05	0.95	0.93
50.00	0.50	0.25	0.95	0.94
50.00	0.50	0.50	0.95	0.94
50.00	0.50	1.00	0.95	0.95
100.00	0.50	0.05	0.95	0.94
100.00	0.50	0.25	0.95	0.95
100.00	0.50	0.50	0.95	0.95
100.00	0.50	1.00	0.95	0.95
200.00	0.50	0.05	0.95	0.94
200.00	0.50	0.25	0.95	0.95
200.00	0.50	0.50	0.95	0.95
200.00	0.50	1.00	0.95	0.95
50.00	2.00	0.11	0.95	0.94
50.00	2.00	0.56	0.95	0.04
50.00	2.00	1.13	0.96	0.94
50.00	2.00	2.26	0.96	0.95
100.00	2.00	0.11	0.95	0.95
100.00	2.00	0.56	0.95	0.00
100.00	2.00	1.13	0.95	0.95
100.00	2.00	2.26	0.97	0.95
200.00	2.00	0.11	0.95	0.95
200.00	2.00	0.56	0.95	0.00
200.00	2.00	1.13	0.95	0.95
200.00	2.00	2.26	0.96	0.95

表 3.9 パラメータの被覆確率, ガンマ分布を仮定した場合

λ	k	β	k の被覆確率	σ の被覆確率
50.00	0.50	0.20	0.95	0.92
50.00	0.50	1.00	0.95	0.93
50.00	0.50	2.00	0.96	0.93
50.00	0.50	4.00	0.95	0.93
100.00	0.50	0.20	0.95	0.93
100.00	0.50	1.00	0.95	0.94
100.00	0.50	2.00	0.95	0.94
100.00	0.50	4.00	0.95	0.94
200.00	0.50	0.20	0.95	0.94
200.00	0.50	1.00	0.95	0.95
200.00	0.50	2.00	0.95	0.95
200.00	0.50	4.00	0.95	0.94
50.00	2.00	0.05	0.95	0.91
50.00	2.00	0.25	0.96	0.92
50.00	2.00	0.50	0.96	0.92
50.00	2.00	1.00	0.97	0.90
100.00	2.00	0.05	0.95	0.93
100.00	2.00	0.25	0.95	0.93
100.00	2.00	0.50	0.95	0.93
100.00	2.00	1.00	0.96	0.93
200.00	2.00	0.05	0.95	0.94
200.00	2.00	0.25	0.95	0.94
200.00	2.00	0.50	0.95	0.94
200.00	2.00	1.00	0.95	0.94
200.00	2.00	1.00	0.95	0.94

第 4 章

窓打切りされた交代再生過程のパラメータの最尤推定

4.1 はじめに

ここでは再発事象のうち、特に交代再生過程としてモデル化できる現象について述べる。窓打切りされた交代再生過程を確率過程として記述し、尤度関数を導出する。また最尤推定量の性質をシミュレーションを用いて明らかにする。この章の内容は阿部・鎌倉 (2016) に基づく。

4.2 窓打切りされた交代再生過程

節の議論に習い、窓打切りされた交代再生過程を構成する。組の変数 (X_i, Y_i) ($i = 1, 2, \dots$) を考える。 $X_2, \dots, X_n, Y_2, \dots, Y_n$ 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_0, μ_1 を持つとする。 (X_1, Y_1) は、以下の分布に従う。

$$\frac{\mu_1}{\mu_1 + \mu_0} F_0(y) \int_0^x \frac{1 - F_1(u)}{\mu_1} du + \frac{\mu_0}{\mu_1 + \mu_0} \int_0^y \frac{1 - F_0(u)}{\mu_0} du$$

我々はこの過程を時刻 w まで観測する (図 4.1)。

4.3 モデル

定期的に故障し、休止期間を経てサービスに復帰するシステムを考える。このシステムは稼働と休止という 2 つの状態を交互に繰り返す。稼働時間と休止時間は相異なる確率分布に従う。このようなシステムは 2.6 で述べた交代再生過程を形作る。我々はイベントの生起を限定された窓を通じてのみ観測する。窓内の最初の観測に関しては直前のイベント生起が確認されず、最後の観測に関しては直後のイベントの生起が観測されない。交代再生過程のサンプルパスに対して窓はただ一つである。

我々の研究の動機となったのはヒートシールの強度分析である。樹脂の素材を溶着させた際に界面がどの程度接着され、どの程度未接着であるか、顕微鏡写真から判断する必要が生じた。これは交互に現れる接着面、未接着面の長さの分布を顕微鏡の視野という限定された窓を通して推定する問題であ

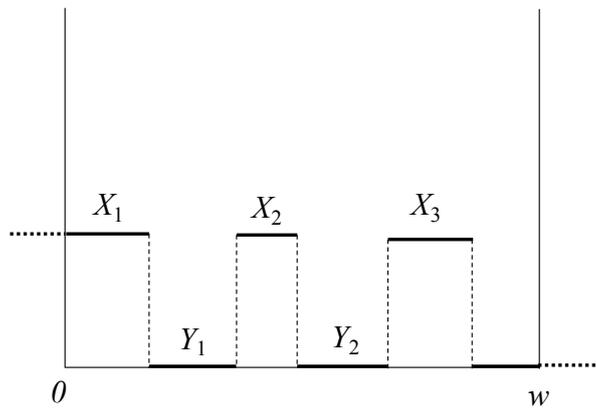


図 4.1 窓打ち切りされた交代再生過程. X_1 の従う分布と X_2 の従う分布は異なる. 窓の外側で起きたイベントについては観測されない.

り, 窓打ち切りされた交代再生過程としてモデル化される (図 4.2). 界面がどの程度の割合で接着され

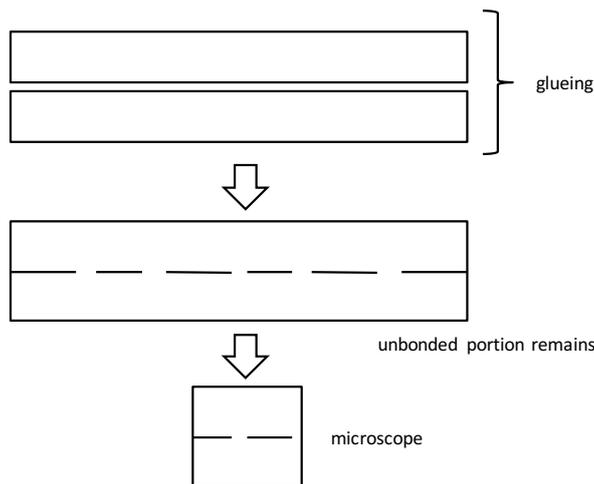


図 4.2 顕微鏡写真の模式図. 黒い線分は接着されていない界面.

ているかという点には特に関心がある. 接着面の占める割合を推定する問題は, 信頼性工学における極限アベイラビリティを推定する問題と同一である.

4.4 尤度関数の導出

複数の窓から得られる観測に基づいた尤度関数を構成する前に, ある一つの窓 i からなる観測を見て部分尤度を構成する.

以降, 状態 1 の長さ X が従う確率分布の密度関数を $f_1(x)$, 分布関数を $F_1(x)$, X の均衡分布の密

密度関数を $g_1(x)$, 均衡分布の生存関数を $G_1(x)$ と書く. 状態 0 の長さ Y についても同様の添字で, 密度関数を $f_0(y)$, 分布関数を $F_0(y)$, Y の均衡分布の密度関数を $g_0(y)$, 均衡分布の生存関数を $G_0(y)$ と表す.

τ_{ij} を窓 i における j 番目の状態 1 または 0 の長さの観測値とする. $s_{i,j}$ を窓 i における j 番目の観測が状態 1 のとき 1, 状態 0 の場合に 0 の値を取る変数とする. はじめに, 観測が状態 1 から始まり, 観測される状態 1 または 0 の長さの総数が 1 のとき, 尤度は,

$$L_{1,1} = \rho G_1(\tau_{i1})$$

である. ここで $\rho = \mu_1/(\mu_0, \mu_1)$ とした.

次に, 観測が状態 1 から始まり, 観測される状態 1 または 0 の長さの総数が 2 のとき, 尤度は,

$$L_{1,2} = \rho g_1(\tau_{i1}) \{1 - F_0(\tau_{i2})\}$$

である.

観測が状態 1 から始まり, 観測される状態 1 または 0 の長さの総数が $n_i \geq 3$ のとき, 尤度は,

$$\begin{aligned} L_{1,n_i} &= \rho g_1(\tau_{i1}) \\ &\times \left(\prod_{j=2}^{n_i-1} \{f_1(\tau_{ij})\}^{s_{i,j}} \{f_0(\tau_{ij})\}^{1-s_{i,j}} \right) \\ &\times \{1 - F_1(\tau_{in_i})\}^{s_{i,n_i}} \{1 - F_0(\tau_{in_i})\}^{1-s_{i,n_i}} \end{aligned}$$

である.

観測が状態 0 から始まる場合の尤度も, 同様にして構成される. 観測が状態 0 から始まり, 観測される状態 1 または 0 の長さの総数が 1 のとき, 尤度は,

$$L_{0,1} = (1 - \rho) G_0(\tau_{i1}).$$

観測が状態 0 から始まり, 観測される状態 1 または 0 の長さの総数が 2 のとき, 尤度は,

$$L_{0,2} = (1 - \rho) g_0(\tau_{i1}) \{1 - F_1(\tau_{i2})\}.$$

観測が状態 0 から始まり, 観測される状態 1 または 0 の長さの総数が $n_i \geq 3$ のとき, 尤度は,

$$\begin{aligned} L_{0,n_i} &= (1 - \rho) g_0(\tau_{i1}) \\ &\times \left(\prod_{j=2}^{n_i-1} \{f_1(\tau_{ij})\}^{s_{i,j}} \{f_0(\tau_{ij})\}^{1-s_{i,j}} \right) \\ &\times \{1 - F_1(\tau_{in_i})\}^{s_{i,n_i}} \{1 - F_0(\tau_{in_i})\}^{1-s_{i,n_i}} \end{aligned}$$

である.

サンプルが複数の窓 $i = 1, \dots, n$ からなる場合, 尤度は,

$$L \propto \prod_{i=1}^n L_{s_{i,1}, n_i} \quad (4.1)$$

と書ける. ここで n_i は窓 i における観測の総数である. この L が目標としていた最大化すべき全尤度である.

この記法は尤度の導出のためには簡単であるが, 実際に最適化するにあたっては便利ではない. そのため, (4.1) の尤度を表現し直す. τ_{ij} が窓ごとに分かれていると尤度を書く際に不便なため, 窓 i を表す添字を落とし, m を観測される状態 1 または 0 の長さの総数とし, $(\tau'_1, \dots, \tau'_m) \stackrel{d}{=} (\tau_{1,1}, \dots, \tau_{n,n_n})$ と改めて置く. 同様に s_{ij} についても, 窓 i を表す添字を落とし, $(s'_1, \dots, s'_m) \stackrel{d}{=} (s_{1,1}, \dots, s_{n,n_n})$ と改めて置く. 対応して a_i を ($i = 1, \dots, m$) を τ'_i が窓の始点に当たる観測の場合に 0, そうでなければ 1 の値を取る変数とする. また d_i ($i = 1, \dots, m$) を窓の終点 (右打切り) に当たる観測の場合に 1, そうでなければ 0 の値を取る変数とする. サンプルサイズ (窓の数) n が与えられたとき, 観測が状態 1 から始まる窓の数 n_1 はパラメータ ρ の二項分布

$$\binom{n}{n_1} \rho^{n_1} (1-\rho)^{n-n_1}$$

に従う. よって (4.1) 式は

$$\begin{aligned} L &\propto \rho^{n_1} (1-\rho)^{n-n_1} \times \\ &\prod_{i=1}^m \left(\left[\{f_1(\tau'_i)\}^{d_i} \{1-F_1(\tau'_i)\}^{1-d_i} \right]^{1-a_i} \right. \\ &\quad \times \left. \left[\{g_1(\tau'_i)\}^{d_i} \{G_1(\tau'_i)\}^{1-d_i} \right]^{a_i} \right)^{s'_i} \\ &\quad \times \left(\left[\{f_0(\tau'_i)\}^{d_i} \{1-F_0(\tau'_i)\}^{1-d_i} \right]^{1-a_i} \right. \\ &\quad \times \left. \left[\{g_0(\tau'_i)\}^{d_i} \{G_0(\tau'_i)\}^{1-d_i} \right]^{a_i} \right)^{1-s'_i} \end{aligned} \quad (4.2)$$

と改めて書き直せる.

この尤度を最大化することでパラメータを求める. 最尤推定量は多くの場合解析的には求まらない. 本研究では準ニュートン法, またはネルダー・ミード法 (Nelder & Mead, 1965) を用いて最適化した.

4.5 尤度関数の左右対称性について

時間的なデータの窓打切りを扱う場合, 窓の始点と終点は明確に区別できる. しかし本研究の動機となったヒートシールの界面のように空間的なデータを扱う場合, 窓の左端を始点とすべきか, 右端を始点とすべきかはわからない. しかし, それでもなお, (4.2) 式に示した尤度を用いることができる.

観測が状態 1 から始まる窓の数は $n_1 = \sum_{i=1}^m (1-a_i)s'_i$ であり, $n-n_1 = \sum_{i=1}^m (1-a_i)(1-s'_i)$

であるから, (4.2) 式は,

$$\begin{aligned} L \propto \prod_{i=1}^m & \left(\left[\{f_1(\tau'_i)\}^{d_i} \{1 - F_1(\tau'_i)\}^{1-d_i} \right]^{a_i} \right. \\ & \times \left. \left[\rho \{g_1(\tau'_i)\}^{d_i} \{G_1(\tau'_i)\}^{1-d_i} \right]^{1-a_i} \right)^{s'_i} \\ & \times \left(\left[\{f_0(\tau'_i)\}^{d_i} \{1 - F_0(\tau'_i)\}^{1-d_i} \right]^{a_i} \right. \\ & \times \left. \left[(1-\rho) \{g_0(\tau'_i)\}^{d_i} \{G_0(\tau'_i)\}^{1-d_i} \right]^{1-a_i} \right)^{1-s'_i} \end{aligned}$$

と書き直せる. さらに, $\rho = \mu_1/(\mu_0 + \mu_1)$, $1 - \rho = \mu_0/(\mu_0 + \mu_1)$, $g_i(\tau) = \{1 - F(\tau)\}/\mu_i$, $G(\tau) = [\int_{\tau}^{\infty} \{1 - F(u)\} dy] / \mu_i$ ($i = 0, 1$) であることに注意すると, (4.2) 式は,

$$\begin{aligned} L \propto \left(\frac{1}{\mu_0 + \mu_1} \right)^n & \prod_{i=1}^m \left(\left[\{f_1(\tau'_i)\}^{d_i} \{1 - F_1(\tau'_i)\}^{1-d_i} \right]^{a_i} \right. \\ & \times \left. \left[\{1 - F_1(\tau'_i)\}^{d_i} \left\{ \int_{\tau'_i}^{\infty} \{1 - F_1(\tau'_i)\} \right\}^{1-d_i} \right]^{1-a_i} \right)^{s'_i} \\ & \times \left(\left[\{f_0(\tau'_i)\}^{d_i} \{1 - F_0(\tau'_i)\}^{1-d_i} \right]^{a_i} \right. \\ & \times \left. \left[\{1 - F_0(\tau'_i)\}^{d_i} \left\{ \int_{\tau'_i}^{\infty} \{1 - F_0(u)\} du \right\}^{1-d_i} \right]^{1-a_i} \right)^{1-s'_i} \quad (4.3) \end{aligned}$$

と表すことができる. 窓の始点と終点を反転させることは, a_i と d_i を入れ替えることに他ならない. (4.3) 式に置いて, a_i と d_i が入れ替わったとしても, 評価される尤度関数はまったく同じである.

4.6 アベイラビリティの簡便な推定量

極限アベイラビリティの簡便な推定量として, 以下を用いることができる.

$$\tilde{\rho} = \left(\sum_{i=1}^m \tau'_i s'_i \right) / (nw) \quad (4.4)$$

これは標本から求めた, 全体の長さに対する状態 1 の長さの総和の比率である.

4.7 シミュレーションによる推定量の比較

本節では (4.4) 式に表したアベイラビリティの簡便な推定量 $\hat{\rho}$ と, 最尤推定量を比較する.

窓打切りされた交代再生過程をシミュレートするアルゴリズムは Algorithm 2 に記載した. このアルゴリズムでは, 密度関数 $f_1(x)$ を持つ分布に従う乱数と, $f_0(x)$ に従う乱数を交互に発生させ, 累

Algorithm 2 窓打ち切りされた交代再生過程をシミュレートするアルゴリズム

Require: 観測開始の原点 T_0 と窓の幅 w および $f_0(x)$, $f_1(x)$ のパラメータ.

Ensure: τ_i, s'_i, a'_i, d'_i , ($i = 1, \dots, m$).

```

while  $\sum_{i=1} t_i < T_0 + w$  do
  generate  $t_{2i-1} \sim F_1(\cdot)$  ( $i = 1, 2, \dots$ ).
  generate  $t_{2i} \sim F_0(\cdot)$ 
end while
 $\sum_{i=1} t_i$  が  $T_0$  を超える最小の  $i$  を  $k$  とする.
 $\sum_{i=1} t_i$  が  $T_0 + w$  を超える最小の  $i$  を  $l$  とする.
if  $k = l$  then
   $\tau'_1 = w, d_1 = 0, a_1 = 1$  とし,  $k \bmod 2$  を  $s'_1$ 
else
   $\tau'_1 = \sum_{i=1}^k t_i - T_0$ 
   $t_{k+1} = \tau'_2, t_{k+2} = \tau'_3, \dots, t_{l-1} = \tau'_{n-1}$  とする.
   $d_1 = 1, d_2 = 1, \dots, d_{n-1} = 1, d_n = 0$  とする.
   $a_1 = 0, a_2 = 1, \dots, a_{n-1} = 1, a_n = 1$  とする.
   $(k, k+1, \dots, l) \bmod 2$  を  $(s'_1, \dots, s'_n)$  とする.
end if

```

積和を取り, 区間 $[0, w]$ の範囲の乱数をデータセットとする. この擬似的なデータセットから $f_1(x)$, $f_0(x)$ のパラメータを推定する.

各状態における状態 1 または 0 の長さの分布には指数分布, ワイブル分布, ガンマ分布を仮定する. 指数分布については状態 1 と 0 でそれぞれ密度関数

$$f_i(x) = \frac{1}{\mu_i} \exp\left(-\frac{x}{\mu_i}\right), \quad i = 0, 1$$

を持つものとした. ワイブル分布については,

$$f_i(x) = \frac{m_i}{\eta_i} \left(\frac{x}{\eta_i}\right)^{m_i-1} \exp\left(-\left(\frac{x}{\eta_i}\right)^{m_i}\right), \quad i = 0, 1,$$

ガンマ分布については,

$$f_i(x) = \frac{1}{\Gamma(k)\sigma_i^k} x^{k-1} e^{-x/\sigma_i}, \quad i = 0, 1,$$

とパラメタライズした.

4.7.1 アベイラビリティの推定

この節におけるシミュレーションの目的は $\tilde{\rho} = (\sum \tau'_i s'_i)/(nw)$ と最尤推定量から求めた $\hat{\rho}$ を比較することである.

サンプルサイズを $n = 10$ で固定し、観測範囲 w を 10 から 100 まで 10 段階で変化させ平均二乗誤差 (MSE) の様子を見る。各 w 毎に 1 万回のシミュレーションを行い結果を図 4.3-4.9 に示した。

$m_1 = 1, m_2 = 1$ で既知のとき、すなわち指数分布の場合、窓の幅が十分に大きければ、簡便法による推定量が最尤推定量にさほど劣らない精度を持つ (図 4.3)。ワイブル分布の場合、MSE の挙動は形状パラメータによって変化が見られる。形状パラメータが $m = 1$ から離れると、最尤法による推定が簡便法に比べ優れた精度を持つ傾向がある (図 4.4-4.9)。ワイブル分布の形状パラメータは、故障現象において意味を持つ。 $m < 1$ ならば、時間とともに故障率が小さくなる性質と解釈でき、 $m > 1$ は、反対に時間とともに故障率が大きくなる性質を示す。シミュレーション結果からは時間に対して故障率が一定でないような事象に対しては、最尤法による推定を行う価値があると解釈できる。

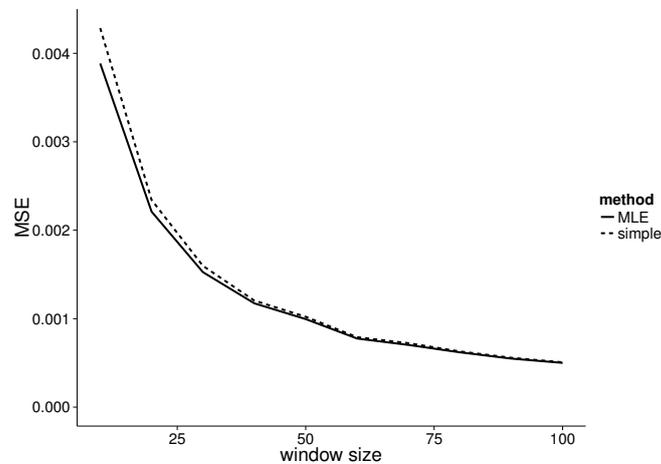


図 4.3 極限アベイラビリティの推定結果. 指数分布を仮定し、パラメータ $\eta_1 = 8, \eta_2 = 2$ のとき。サンプルサイズ 10 のとき。

これらの傾向は小標本、大標本の場合も同様である。再生間隔の分布に指数分布を仮定し、サンプルサイズ 5 の場合が図 4.10、サンプルサイズ 100 の場合が図 4.11 である。ワイブル分布を仮定し、サンプルサイズを 5 とした場合の結果を図 4.12 - 4.13 に、サンプルサイズを 100 とした場合の結果を図 4.14 - 4.15 に示した。

信頼性工学の分野ではワイブル分布がよく用いられるが、ガンマ分布もまた指数分布の拡張の一つである。ガンマ分布を仮定した場合の、シミュレーション結果を図 4.16-4.18 に示す。ワイブル分布を仮定した場合ほど大きな差ではないが MSE の改善が見られる。

4.8 この章のまとめ

この章では窓打ち切りされた交代再生過程に対応する尤度を導出し、窓打ち切りされた交代再生過程のパラメータ推定を可能にした。4.7.1 節で評価した通り、本手法はアベイラビリティの推定精度の改善に役立てることができる。一方で、アベイラビリティのみを推定するものではなく、交代再生過程の再

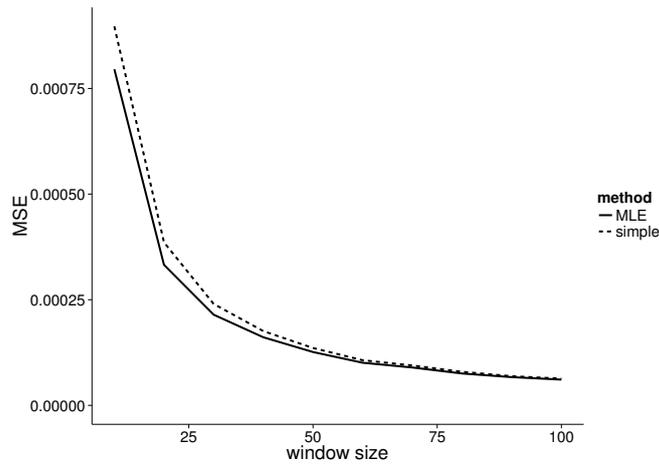


図 4.4 極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8, \eta_2 = 2, m_1 = 3, m_2 = 3$, サンプルサイズ 10 のとき.

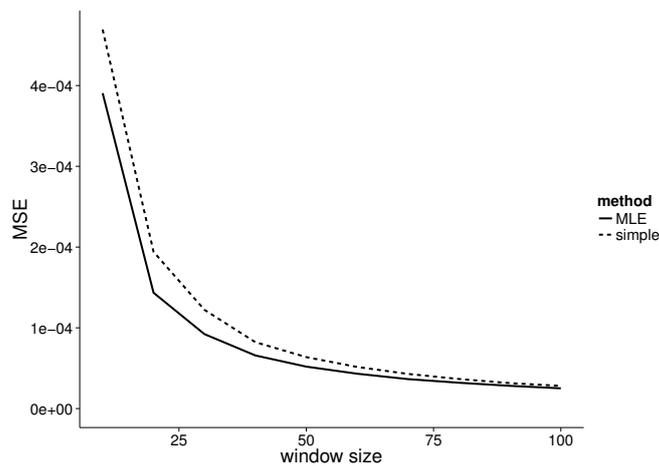


図 4.5 極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8, \eta_2 = 2, m_1 = 5, m_2 = 5$, サンプルサイズ 10 のとき.

生間隔の分布を推定するものである. ヒートシールの強度を分析する際, 短い未接着面が多くある場合と, 長い未接着面が少なくある場合では, 比率は同じでも強度に与える影響は異なることが考えられるため, 分布についての情報が得られることは重要である.

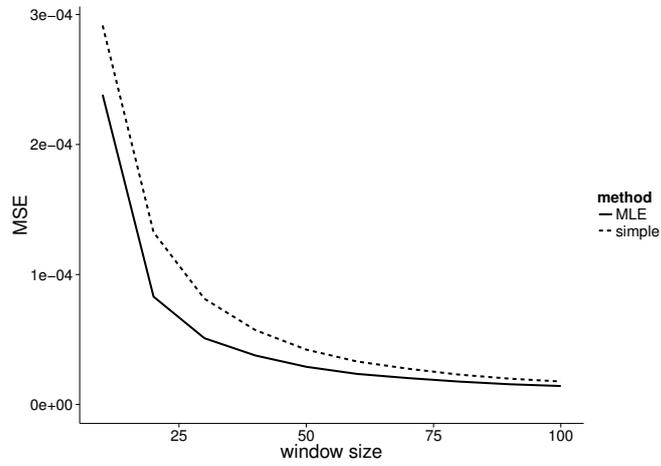


図 4.6 極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8, \eta_2 = 2, m_1 = 7, m_2 = 7$, サンプルサイズ 10 のとき.

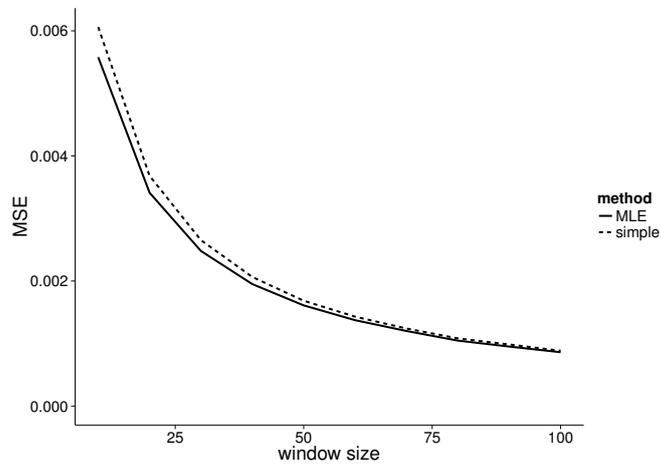


図 4.7 極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8, \eta_2 = 2, m_1 = 0.8, m_2 = 0.8$, サンプルサイズ 10 のとき.

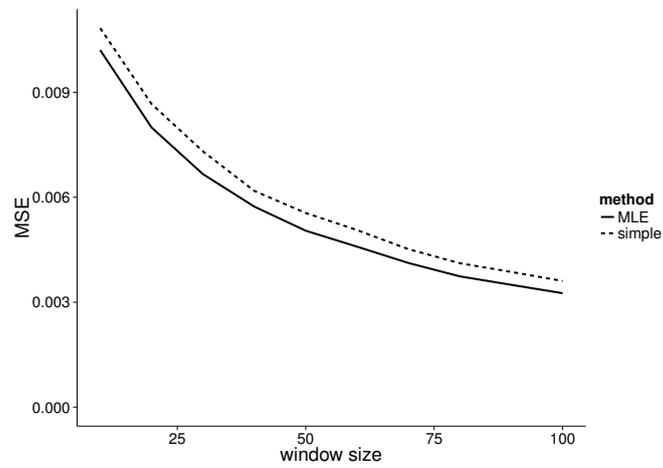


図 4.8 極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8, \eta_2 = 2, m_1 = 0.5, m_2 = 0.5$, サンプルサイズ 10 のとき.

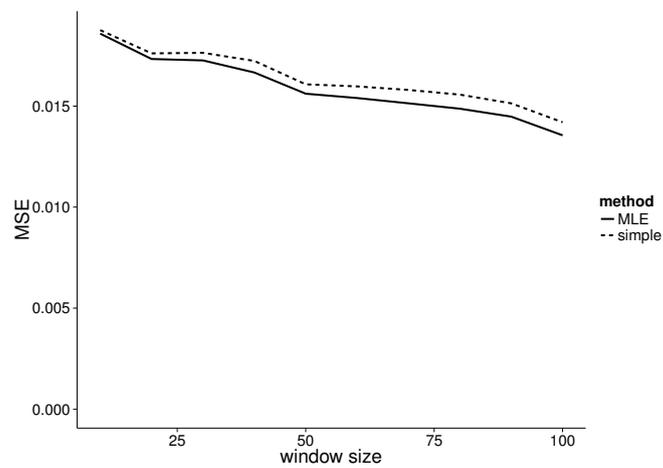


図 4.9 極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8, \eta_2 = 2, m_1 = 0.3, m_2 = 0.3$, サンプルサイズ 10 のとき.

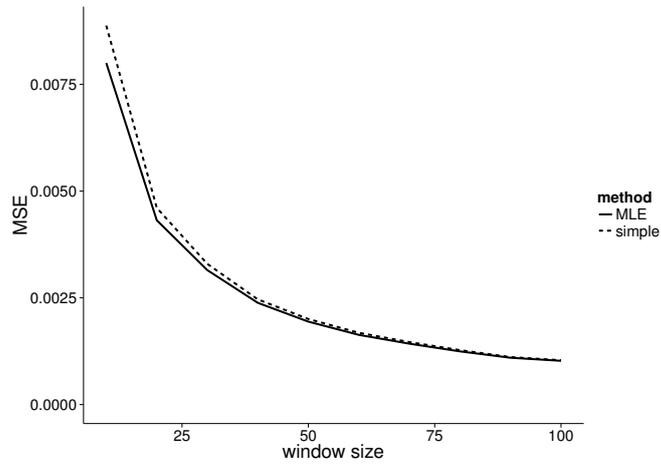


図 4.10 極限アベイラビリティの推定結果. 指数分布を仮定し, パラメータ $\eta_1 = 8, \eta_2 = 2$, サンプルサイズ 5 のとき.

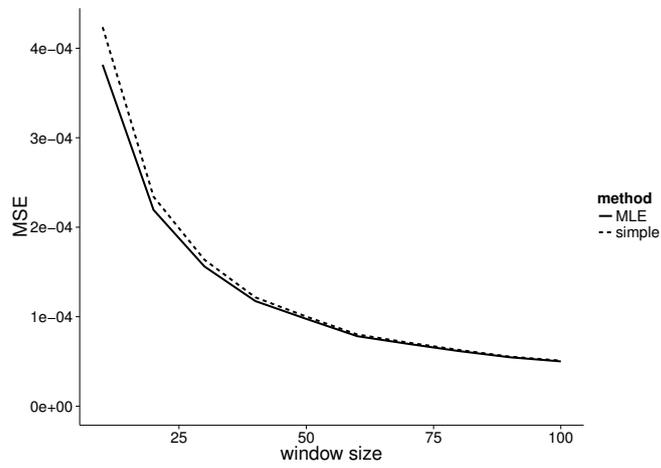


図 4.11 極限アベイラビリティの推定結果. 指数分布を仮定し, パラメータ $\eta_1 = 8, \eta_2 = 2$, サンプルサイズ 100 のとき.

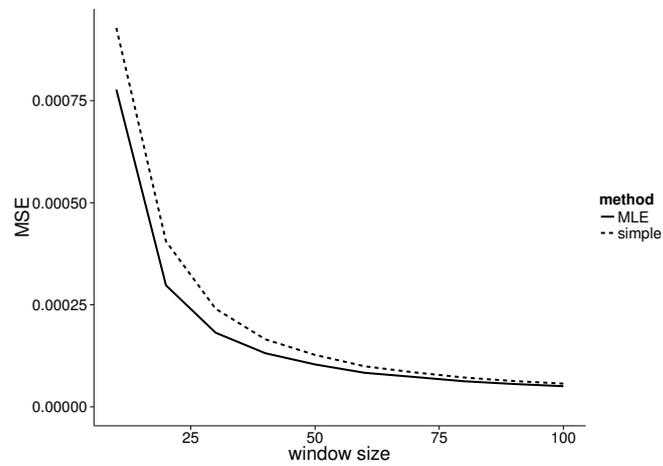


図 4.12 極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8$, $\eta_2 = 2, m_1 = 5, m_2 = 5$, サンプルサイズ 5 のとき.

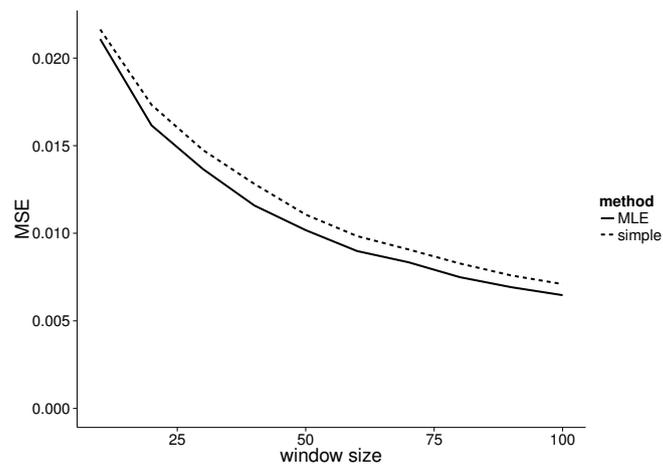


図 4.13 極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8$, $\eta_2 = 2, m_1 = 0.5, m_2 = 0.5$, サンプルサイズ 5 のとき.

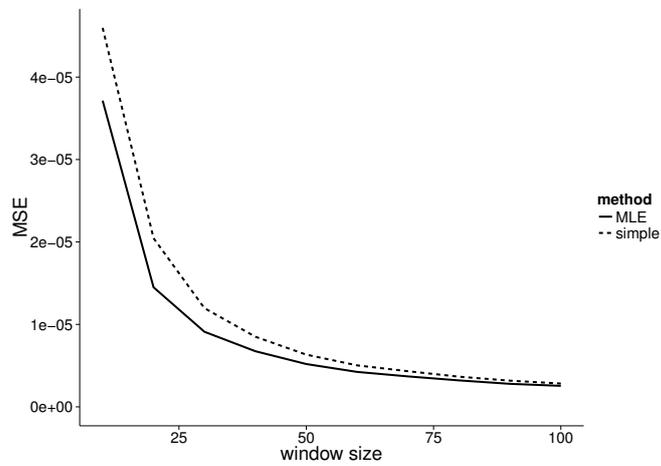


図 4.14 極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8$, $\eta_2 = 2, m_1 = 5, m_2 = 5$, サンプルサイズ 100 のとき.

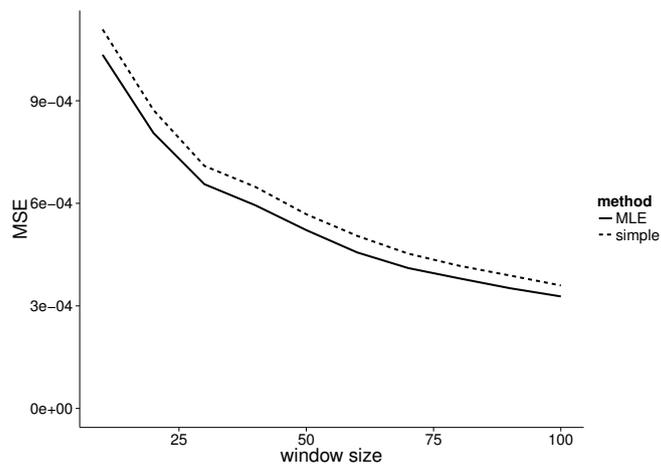


図 4.15 極限アベイラビリティの推定結果. ワイブル分布を仮定し, パラメータ $\eta_1 = 8$, $\eta_2 = 2, m_1 = 0.5, m_2 = 0.5$, サンプルサイズ 100 のとき.

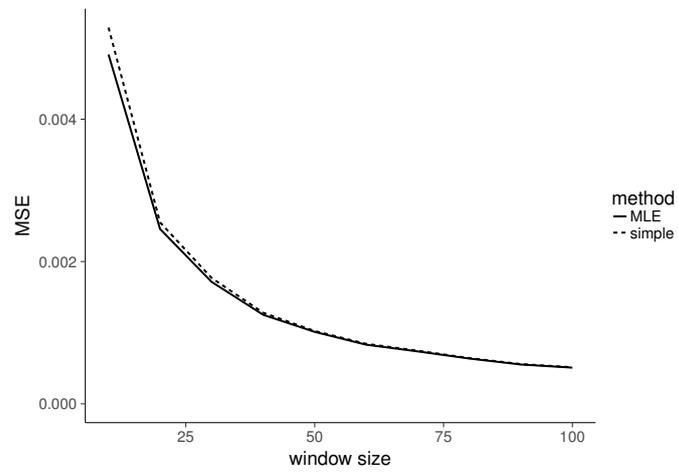


図 4.16 極限アベイラビリティの推定結果. ガンマ分布を仮定し, パラメータ $k_1 = 2, k_2 = 2, \sigma_1 = 4, \sigma_2 = 1$, サンプルサイズ 5 のとき.

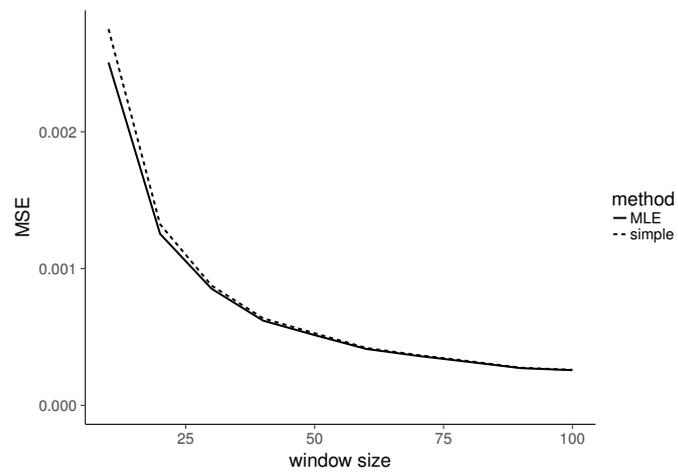


図 4.17 極限アベイラビリティの推定結果. ガンマ分布を仮定し, パラメータ $k_1 = 2, k_2 = 2, \sigma_1 = 4, \sigma_2 = 1$, サンプルサイズ 10 のとき.

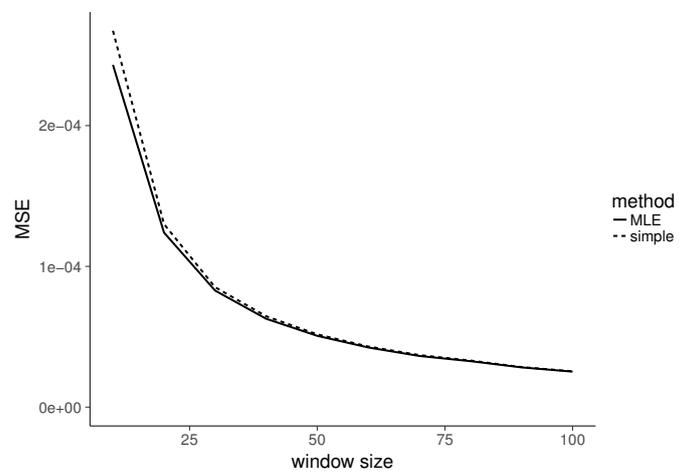


図 4.18 極限アベイラビリティの推定結果. ガンマ分布を仮定し, パラメータ $k_1 = 2, k_2 = 2, \sigma_1 = 4, \sigma_2 = 1$, サンプルサイズ 100 のとき.

第 5 章

窓打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

5.1 はじめに

4 章では阿部・鎌倉 (2016) に基づき, 窓打切りされた交代再生過程の尤度関数の導出について述べた. また, 最尤推定量がアベイラビリティの推定量として MSE の面で優れていることを, シミュレーションで検証した.

阿部・鎌倉 (2016) と同時期に Rootzén & Zholud (2016) でも, 同様の尤度関数を導出している. しかし, Rootzén & Zholud(2016) では, 条件付き最尤法に基づく推定を行っている. この章ではこれら 2 つの推定量を比較する.

また交代再生過程を用いた回帰型のモデルによって, ヒートシールのデータに関する分析を行う.

この章の内容は阿部・鎌倉 (2017) に基づく.

なお, 以降では表記が煩雑になるのを避けるため, 前節の記号を, $(\tau_1, \dots, \tau_m) = (\tau'_1, \dots, \tau'_m)$, $(s_1, \dots, s_m) = (s'_1, \dots, s'_m)$ と改めて置く.

5.2 打切りのメカニズム

ここではまず Rootzén & Zholud (2016) が扱った問題の打切りのメカニズムを整理したい. ここでは二種類の打切りが混在したような状況が考えられた. まず一回の観測で状態 0, 1 の長さの合計が窓の幅 w を超えたら打ち切られる. また, 一回の観測で状態 0 の終了を観測したらその時点で観測を打ち切る. すなわち観測には以下の 5 種類があり得る.

- nc 観測が状態 1 から始まり, 続く状態 0 が完全に観測される場合, その事象の生起する密度関数は $g_1(\tau_1)f_0(\tau_2)$ である (non-censored).
- rc 観測が状態 1 から始まり, 続く状態 0 が右側で打ち切られる場合. 尤度は $g_1(\tau_1)\{1-F_0(\tau_2)\}$ である (right-censored).
- lc 観測が状態 0 から始まり, 状態 0 の終了まで観測されたとき, 密度関数は $g_0(\tau)$ である

(left-censored).

dc 観測が状態 0 から始まり, かつ状態 0 が右側で打ち切られる場合, 尤度は $G_0(w)$ である (doubly-censored).

no 状態 0 が観測されないとき, 尤度は $G_1(w)$ である (no-observed).

状態 0 が観測されない場合 (観測 no) については Rootzén & Zholud (2016) では扱いが明記されていないが, 現実的には起こり得るので, これも付け加えることにする. この 5 種類の打ち切りを図 5.1 で示した. このような打ち切りを Rootzén & Zholud タイプの打ち切りと呼ぶことにする.

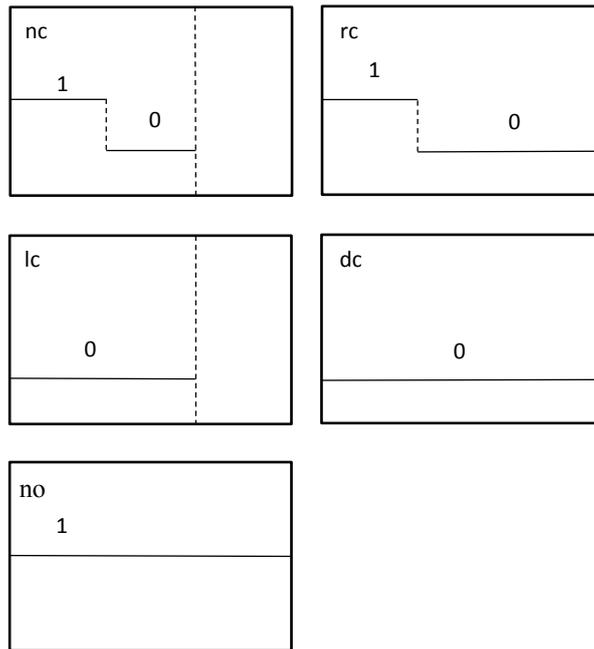


図 5.1 Rootzén & Zholud (2016) が扱った 5 種類の観測.

一方, 阿部・鎌倉 (2016) では, 一回の観測で状態 0, 1 の長さの合計が窓の幅 w を超えた場合にのみ打ち切られ, 一つの窓に対して, 複数回状態 0, 状態 1 が起こりうる. 以降, 単に窓打ち切りと呼ぶ場合は, この打ち切りを指すものとする.

5.3 条件付き最尤法に基づく推定

Rootzén (2016) では, (4.2) 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,

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

を用いた. ここで θ_0 は, $f_0(x)$ 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 (4.2) 式の最初の $\rho^{n_1}(1-\rho)^{n-n_1}$ の因子が消える. 条件付けした理由については明示していないが, 状態 0 の長さの分布にのみ関心があり, 状態 1 には関心がないため, または事例研究で一般化パレート分布を使用しており, ある閾値を超えた長さの分布のみをモデル化する上で, 全体の長さの母平均で定義される ρ の因子が邪魔になるからである. 以降, (5.1) 式を最大化する推定量を求める方法を, Rootzén & Zholud (RZ) 法と呼ぶことにする.

一方で阿部・鎌倉 (2016) では全尤度関数 (4.2) 式を用いて, $f_1(x)$ に関するパラメータ θ_1 と, $f_0(x)$ に関するパラメータ θ_0 を同時に求めることを提案している. 以降, この方法を阿部・鎌倉 (2016) の方法と呼ぶ. 我々は ρ の推定に関心があった. この量は信頼性工学の分野でアベイラビリティと呼ばれる. 以下に続く節で, この 2 種類の推定量を比較する. 状態 0 の長さの分布のみに関心があり, 状態 1 と状態 0 の長さの分布が独立であっても, 阿部・鎌倉 (2016) の方法がバイアスと標準誤差で評価して良い性質を持つことを示す.

5.3.1 指数分布の場合

$F_0(x)$ が指数分布 $F_0(x) = 1 - e^{-x/\mu_0}$ のとき, 均衡分布は元の分布と一致することから, 条件付き対数尤度関数は,

$$l(\mu_0) = -\mu_0^{-1} \sum_{i=1}^m \{(1-s_i)\tau_i\} - \log \mu_0 \sum_{i=1}^m (1-s_i)d_i \quad (5.2)$$

となる. Rootzén & Zholud 法による推定量は

$$\hat{\mu}_0 = \left\{ \sum_{i=1}^m (1-s_i)\tau_i \right\} / \left\{ \sum_{i=1}^m (1-s_i)d_i \right\} \quad (5.3)$$

である.

一方, 全尤度関数の対数は,

$$\begin{aligned}
 l^*(\mu_0, \mu_1) &= n_1 \{ \log \mu_1 - \log(\mu_0 + \mu_1) \} \\
 &\quad + (n - n_1) \{ \log \mu_0 - \log(\mu_0 + \mu_1) \} \\
 &\quad + \sum_{i=1}^m [-s_i d_i \log \mu_1 - \mu_1^{-1} (s_i \tau_i) \\
 &\quad \quad (1 - s_i) d_i \log \mu_0 - \mu_0^{-1} (1 - s_i) \tau_i]
 \end{aligned} \tag{5.4}$$

となる. 阿部・鎌倉 (2016) の方法の推定量は閉じた形では求まらないため, (5.4) 式を数値的に最大化する必要がある.

5.4 推定量の漸近的性質

ここでは指数分布を仮定して, 窓打切りの状況における推定量の性質を調べる.

(5.4) 式を最大化する μ_0, μ_1 を $\tilde{\mu}_0, \tilde{\mu}_1$ と表記する. $\tilde{\mu}_0$ の分散は漸的に,

$$\begin{aligned}
 &\frac{\mu_0^2(\mu_0 + \mu_1)}{nw} \\
 &\times [1 + \mu_1 \mu_0 / \{w(\mu_0 + \mu_1) + \mu_1 \mu_0\}]^{-1}
 \end{aligned} \tag{5.5}$$

となる.

フィッシャー情報行列の負の逆行列を取ることで, 漸近分散を求める. まず式 (5.4) 式の 2 階微分を求める. 1 階微分は,

$$\begin{aligned}
 \frac{\partial \log L(\mu_0, \mu_1)}{\partial \mu_1} &= \frac{l}{\mu_1} - \frac{n}{\mu_0 + \mu_1} - \sum_{i=1}^m \left(\frac{s_i d_i}{\mu_1} - \frac{s_i z_i}{\mu_1^2} \right) \\
 \frac{\partial \log L(\mu_0, \mu_1)}{\partial \mu_0} &= \frac{n - l}{\mu_1} - \frac{n}{\mu_0 + \mu_1} - \sum_{i=1}^m \left(\frac{(1 - s_i) d_i}{\mu_1} - \frac{(1 - s_i) z_i}{\mu_1^2} \right).
 \end{aligned}$$

2 階微分は,

$$\begin{aligned}
 \frac{\partial^2 l^*(\mu_0, \mu_1)}{\partial \mu_1^2} &= \frac{n_1}{\mu_1^2} - \frac{n}{(\mu_0 + \mu_1)^2} \\
 &\quad - \sum_{i=1}^m \left(-\frac{s_i d_i}{\mu_1^2} + 2 \frac{s_i \tau_i}{\mu_1^3} \right).
 \end{aligned}$$

$$\begin{aligned}
 &\frac{\partial^2 l^*(\mu_0, \mu_1)}{\partial \mu_0^2} \\
 &= \frac{n - n_1}{\mu_1^2} - \frac{n}{(\mu_0 + \mu_1)^2} \\
 &\quad - \sum_{i=1}^m \left(-\frac{(1 - s_i) d_i}{\mu_0^2} + 2 \frac{(1 - s_i) \tau_i}{\mu_0^3} \right).
 \end{aligned}$$

$$\frac{\partial^2 l^*(\mu_0, \mu_1)}{\partial \mu_0 \partial \mu_1} = \frac{n}{(\mu_0 + \mu_1)^2}.$$

ここで以下の性質を用いる.

- $E[n_1] = (n\mu_1)/(\mu_0 + \mu_1)$,
- $E[n - n_1] = (n\mu_0)/(\mu_0 + \mu_1)$,
- $E[\sum s_i \tau_i] = (nw\mu_1)/(\mu_0 + \mu_1)$,
- $E[\sum (1 - s_i) \tau_i] = (nw\mu_0)/(\mu_0 + \mu_1)$,
- $E[\sum s_i d_i] = (nw)/(\mu_0 + \mu_1)$,
- $E[\sum (1 - s_i) d_i] = (nw)/(\mu_0 + \mu_1)$.

$$\begin{aligned} I^*(\theta) &= E\left[-\frac{\partial^2}{\partial \theta \partial \theta'} l^*(\mu_0, \mu_1)\right] \\ &= \frac{n}{(\mu_0 + \mu_1)^2} \begin{pmatrix} A & 1 \\ 1 & B \end{pmatrix} \end{aligned}$$

ここで,

$$A = -\frac{1}{\mu_1^2} \{w(\mu_0 + \mu_1) + \mu_1(\mu_0 + \mu_1) - \mu_1^2\},$$

$$B = -\frac{1}{\mu_0^2} \{w(\mu_0 + \mu_1) + \mu_0(\mu_0 + \mu_1) - \mu_0^2\},$$

である. 逆行列を求め,

$$I^*(\theta)^{-1} = \begin{pmatrix} A_2 & C \\ C & B_2 \end{pmatrix}. \quad (5.6)$$

ここで,

$$A_2 = -\frac{\mu_1^2(\mu_0 + \mu_1)}{nw} \frac{1}{1 + \mu_0\mu_1/(w(\mu_0 + \mu_1) + \mu_1\mu_0)},$$

$$B_2 = -\frac{\mu_0^2(\mu_0 + \mu_1)}{nw} \frac{1}{1 + \mu_0\mu_1/(w(\mu_0 + \mu_1) + \mu_1\mu_0)},$$

$$C = -\frac{\mu_0^2\mu_1^2(\mu_0 + \mu_1)}{nw} \frac{1}{w(\mu_0 + \mu_1) + 2\mu_0\mu_1},$$

である. 式 (5.7) に示した事実が求まった.

Rootzén & Zholud 法の推定量 (5.3) 式については, 分散は漸近的に

$$\mu_0^2(\mu_0 + \mu_1)/nw \quad (5.7)$$

となることを示す. これについては Hamilton (1994) 14 章に与えられた, 以下の結果を用いる.

定理 4.

$$E[h(\theta_0, w_t)] = 0$$

とする. ここで w_t はある日 t に観測された厳密に定常なベクトル, θ_0 は未知パラメータの真値の $(a \times 1)$ ベクトル, $h(\cdot)$ は微分可能な r 次元ベクトル値関数で $r > a$. GMM 推定値 $\hat{\theta}_T$ は以下を最小化する θ の値である.

$$[g(\theta; y_T)]' \hat{S}_T^{-1} [g(\theta; y_T)].$$

ここで

$$[g(\theta; y_T)] \equiv (1/T) \sum_{t=1}^T h(\theta, w_t)$$

\hat{S}_T は

$$S = \lim_{T \rightarrow \infty} (1/T) \sum_{t=1}^T \sum_{v=-\infty}^{\infty} E\{[h(\theta_0, w_t)][h(\theta_0, w_{t-v})]'\}$$

の推定値. GMM 推定値は

$$\hat{\theta}_T \approx N(\theta_0, \hat{V}_T/T),$$

ここで

$$\hat{V}_T = \{\hat{D}_T \cdot \hat{S}_T^{-1} \hat{D}_T'\}^{-1},$$

また

$$\hat{D}_T' = \frac{\partial g(\theta; y_T)}{\partial \theta'} \Big|_{\theta=\hat{\theta}_T}.$$

上記の定理を用いるために, まず式 (5.2) 式の導関数を求める.

$$\frac{\partial}{\partial \mu_0} l(\mu_0, \mu_1) = \sum_{i=1}^m \left[-\frac{(1-s_i)d_i}{\mu_0} + \frac{(1-s_i)\tau_i}{\mu_0^2} \right] \quad (5.8)$$

これは,

$$E \left[\frac{\partial}{\partial \mu_0} l(\mu_0) \right] = 0 \quad (5.9)$$

を満たす.

いま,

$$\begin{aligned} E \left\{ \left[\frac{\partial}{\partial \mu_0} l(\mu_0) \right] \left[\frac{\partial}{\partial \mu_0} l(\mu_0) \right] \right\} \\ = -E \left\{ \frac{\partial^2}{\partial \mu_0^2} l(\mu_0) \right\} \end{aligned}$$

であるから条件付き対数尤度関数の2階導関数の期待値に対して、負の逆数を取ると漸近分散が求まる.

2階微分を求めると,

$$\frac{\partial^2}{\partial \mu_0^2} l(\mu_0) = \sum_{i=1}^m \left[\frac{(1-s_i)d_i}{\mu_0^2} - 2 \frac{(1-s_i)z_i}{\mu_0^3} \right].$$

期待値を取り,

$$E \left[\frac{\partial^2}{\partial \mu_0^2} l(\mu_0) \right] = -\frac{nw}{\mu_0^2(\mu_0 + \mu_1)}.$$

5.4節に示した事実が求まった.

以上の議論により、以下の(5.10)式を得る. 阿部・鎌倉(2016)の方法のRootzén & Zholud法に対する漸近相対効率は,

$$1 + \mu_1\mu_0 / \{w(\mu_0 + \mu_1) + \mu_1\mu_0\} > 1 \quad (5.10)$$

となる. すべてのパラメータが正であることに注意して、漸近相対効率は1以上であることがわかる.

5.5 シミュレーションによる推定量の比較

漸近相対効率だけでなく、バイアス、標準誤差を比較するためにシミュレーションを行う. Rootzén & Zholud タイプの打切りと窓打切りの状況をモンテカルロ・シミュレーションで再現するための方法を以下に記載する. (2.34)式の通り、 ρ は状態0の長さの分布と、状態1の長さの分布の平均で決まる量であり、シミュレーションにあたって任意に設定するものではないことに注意する. Rootzén & Zholud タイプの打切りを再現するシミュレーションのアルゴリズムはAlgorithm 3に示した. また、窓打切りの状況を再現するアルゴリズムはAlgorithm 4に記載した. Algorithm 4はRootzén & Zholud (2016)が用いた方法と同じである. 4.7節に示したAlgorithm 2でも窓打切りの状況のシミュレーションは可能だが、Rootzén & Zholud (2016)との比較のため、この節ではAlgorithm 4を用いる.

Algorithm 3 Rootzén & Zholud タイプの打ち切りをシミュレートするアルゴリズム

Require: w および $f_1(x)$, $f_0(x)$ のパラメータ.

Ensure: τ_i, s_i, a_i, d_i , ($i = 1, 2$).

 generate $u \sim \text{uniform}(0, 1)$
 $\rho = \mu_1 / (\mu_0 + \mu_1)$
if $u \leq \rho$ **then**

 generate t_1 with density $g_1(\cdot)$
if $t_1 \leq w$ **then**

 generate t_2 with density $f_0(\cdot)$
if $t_1 + t_2 \leq w$ **then**
 $\tau_1 = t_1, \tau_2 = t_2, s_1 = 1, s_2 = 0, a_1 = 0, a_2 = 1, d_1 = 1, d_0 = 0$
else
 $\tau_1 = t_1, \tau_2 = w - t_1, s_1 = 1, s_2 = 0, a_1 = 0, a_2 = 1, d_1 = 1, d_0 = 0$
end if
else
 $\tau_1 = w, s_1 = 1, a_1 = 0, d_1 = 0$
end if
else

 generate t_1 with density $g_0(\cdot)$
if $t_1 \leq w$ **then**
 $\tau_1 = t_1, s_1 = 0, a_1 = 0, d_1 = 1$
end if
else
 $\tau_1 = w, s_1 = 0, a_1 = 0, d_1 = 0$
end if

密度関数 $f_1(x)$ を持つ分布に従う乱数と、 $f_0(x)$ に従う乱数を交互に発生させ、累積和を取り、区間 $[0, w]$ の範囲の乱数をデータセットとする。この擬似的なデータセットから $f_1(x)$, $f_0(x)$ のパラメータを推定する。

前節に引き続き状態 0 の長さの分布には指数分布、ワイブル分布、ガンマ分布を仮定する。指数分布については状態 1 と 0 でそれぞれ密度関数

$$f_1(x) = \frac{1}{\eta_1} \exp\left(-\frac{x}{\eta_1}\right),$$

$$f_0(y) = \frac{1}{\eta_0} \exp\left(-\frac{y}{\eta_0}\right)$$

Algorithm 4 窓打切りされた交代再生過程をシミュレートするアルゴリズム

Require: w および $f_1(\cdot)$, $f_0(\cdot)$ のパラメータ.

Ensure: τ_i, s_i, a_i, d_i , ($i = 1, \dots, m$).

```

generate  $u \sim \text{uniform}(0, 1)$ 
 $\rho = \mu_1 / (\mu_0 + \mu_1)$ 
if  $u \leq \rho$  then
  generate  $t_1$  with density  $g_1(\cdot)$ 
  generate  $t_{2i}$  with density  $f_0(\cdot)$  ( $i = 1, 2, \dots$ )
  generate  $t_{2i+1}$  with density  $f_1(\cdot)$ 
else
  generate  $t_1$  with density  $g_0(\cdot)$ 
  generate  $t_{2i}$  with density  $f_1(\cdot)$  ( $i = 1, 2, \dots$ )
  generate  $t_{2i+1}$  with density  $f_0(\cdot)$ 
end if
 $l = \min\{i \mid \sum t_i > w\}$ 
 $\tau_l = w - \sum_{i=1}^{l-1} t_i$ 
 $\tau_1 = t_1, \tau_2 = t_2, \tau_3 = t_3, \dots, \tau_{l-1} = t_{l-1}$ 
 $d_1 = 1, d_2 = 1, \dots, d_{l-1} = 1, d_l = 0$ 
 $a_1 = 0, a_2 = 1, \dots, a_{l-1} = 1, a_l = 1$ 
if  $u \leq \rho$  then
   $s_1 = 1 \bmod 2, s_2 = 2 \bmod 2, \dots, s_l = l \bmod 2$ 
else
   $s_1 = 2 \bmod 2, s_2 = 3 \bmod 2, \dots, s_l = l \bmod 2$ 
end if

```

を持つものとした. 状態 0 のワイブル分布については,

$$f_0(x) = \frac{m}{\eta} \left(\frac{x}{\eta}\right)^{m-1} \exp\left(-\left(\frac{x}{\eta}\right)^m\right),$$

とパラメタライズした.

5.5.1 状態 0 の長さの分布のパラメータの推定

本節を通じて, 観測窓の幅は $w = 6$ に設定した. 状態 1 の長さの分布は平均 6 の指数分布とした. 状態 0 の長さの分布は平均 $\mu_0 = 0.2, 1.5$ の指数分布, ワイブル分布, ガンマ分布を選んだ. 状態 1 の長さの分布は常に指数分布とした. Rootzén & Zholud (2016) との比較のため, これらの設定は Rootzén & Zholud (2016) と合わせている. シミュレーションの試行回数は各 10000 回とした. ワイブル分布のシミュレーションにあたっては形状パラメータ $m = 0.5, 2$ とした. 平均

$\mu_0 = \eta_0 \Gamma(1 + 1/m_0)$ を 0.2, 1, 5 とするため, $m = 0.5$ のときは $\eta = 0.1, 0.5, 2.5$ とし, $m = 2$ のときは, $\eta = 0.226, 1.128, 5.642$ とした. ガンマ分布については形状パラメータ $k = 0.5, 3$ とし, 平均 $\mu_0 = k\sigma$ を 0.2, 1, 5 とするため, $k = 0.5$ のときは, $\sigma = 0.4, 2, 10$ とし, $m = 3$ のときは $\sigma = 0.067, 0.333, 1.667$ とした. (4.2) 式, (5.1) 式の関数の最適化には準ニュートン法を用い, 収束しなかった場合に限りネルダー・ミード法 (Nelder & Mead, 1965) を用いた.

本節の図はすべて, Rootzén & Zholud と阿部・鎌倉 (2016) の方法を比較したものであり, 点が推定量の平均, エラーバーが ± 1 標準誤差を表す. 横軸はサンプルサイズ n を表し, また実線のエラーバーが阿部・鎌倉 (2016) の方法, 点線のエラーバーが Rootzén & Zholud 法を表す. 図の上につけられたサブタイトルは, 状態 0 の長さの平均を表示している. 水平の点線はシミュレーションで設定したパラメータの真値である.

指数分布の場合について, まず Rootzén & Zholud タイプの打切りでの推定量の性質を調べる. Rootzén & Zholud 法と阿部・鎌倉 (2016) の方法を比較した結果を図 5.2 に示す. 次に窓打切り状況下で, 指数分布を仮定して, Rootzén & Zholud 法と阿部・鎌倉 (2016) の方法の推定量の性質を比較した結果を図 5.3 に示す.

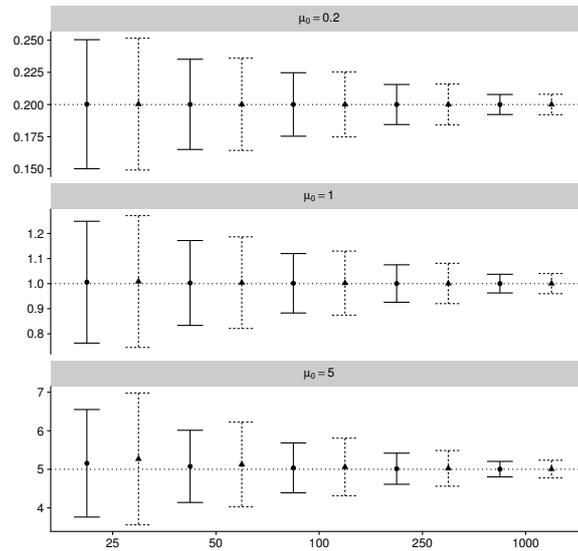


図 5.2 μ_0 の推定値. 指数分布を仮定し, Rootzén & Zhould タイプの打切りの場合.

図 5.2 および図 5.3 を見ると, μ_0 が大きくサンプルサイズが小さい場合, Rootzén & Zholud 法に比べて阿部・鎌倉 (2016) の方法が標準誤差, バイアスともに小さいことがわかる. μ_0 が小さい場合, 3 節で述べた観測 nc (図 5.1) が多くなりやすい. nc が多いとき, またはサンプルサイズが非常に多い場合は, 両者の差は小さくなる.

また, 推定量の分散を窓打切り状況下で比較した. シミュレーションで求めた分散の比率と 5.4 節で求めた漸近相対効率を並べて表示したものが, 表 5.1 である. 表中の ∞ の行は漸近相対効率である. ここでの比率は Rootzén & Zholud 法の推定量の分散を阿部・鎌倉 (2016) の方法の推定量の分

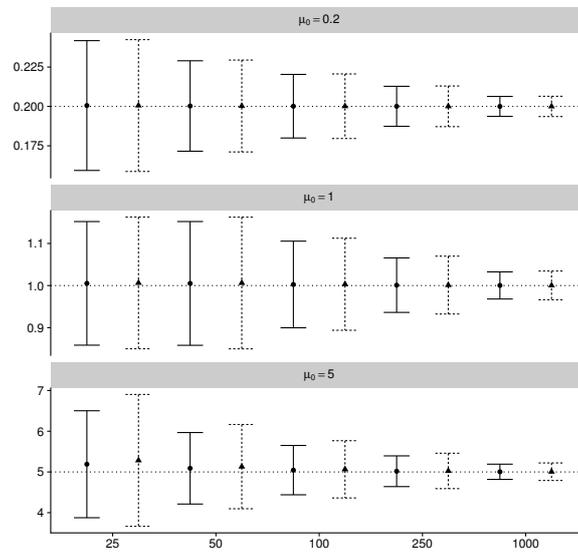


図 5.3 μ_0 の推定値. 指数分布を仮定し, 窓打ちの場合.

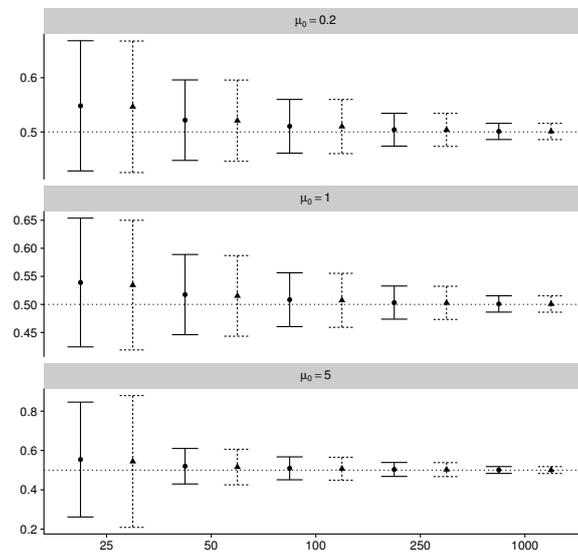


図 5.4 形状パラメータ m の推定値. ワイブル分布 ($m = 0.5$) を仮定し, Rootzén & Zhould タイプの打ちの場合.

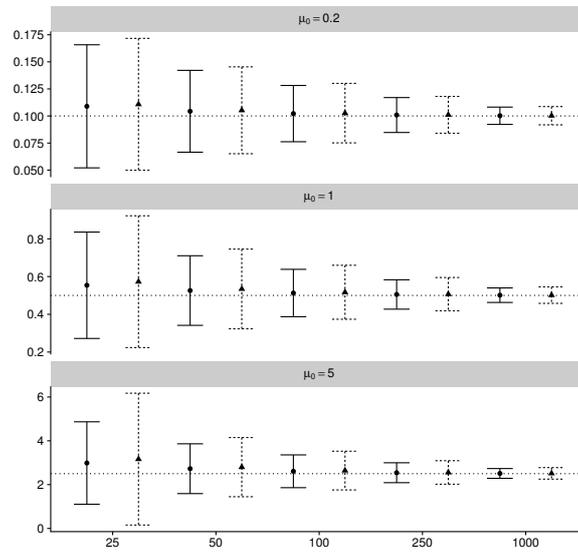


図 5.5 尺度パラメータ η の推定値. ワイブル分布 ($m = 0.5$) を仮定し, Rootzén & Zhould タイプの打切りの場合.

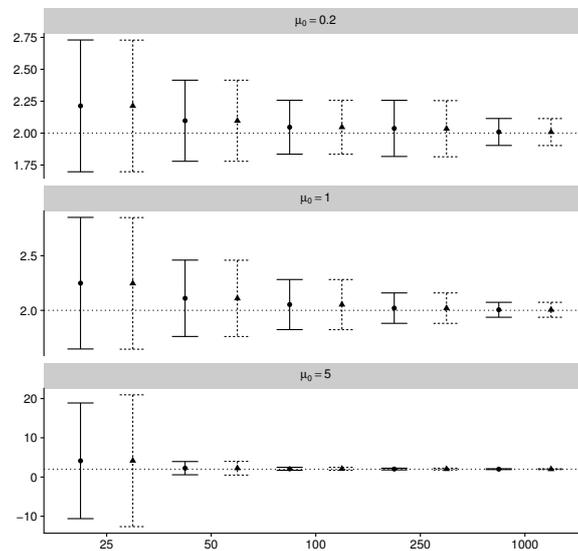


図 5.6 形状パラメータ m の推定値. ワイブル分布 ($m = 2$) を仮定し, Rootzén & Zhould タイプの打切りの場合.

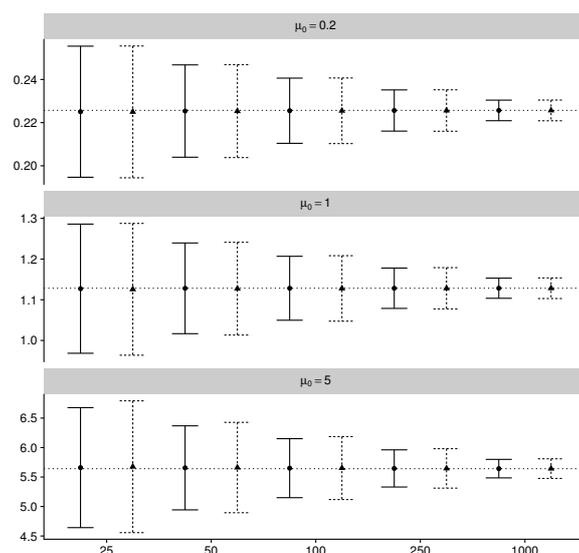


図 5.7 尺度パラメータ η の推定値. ワイブル分布 ($m = 2$) を仮定し, Rootzén & Zholud タイプの打ち切りの場合,

散で割ったものである. その値は 1 を超えており, 阿部・鎌倉 (2016) の方法のほうが効率が良いことを示している. この傾向は μ_0 が大きくサンプルサイズが小さい場合特に顕著である. 最尤推定量は漸近的に 5.4 節に示した分散を持つ. 小標本における性質はシミュレーション (図 5.3) により示した. n が小さい場合の標準誤差は阿部・鎌倉 (2016) の方法のほうが小さく保たれるため, この傾向が現れる.

Rootzén & Zholud タイプの打ち切りで, ワイブル分布を仮定した場合については図 5.4 – 5.7 に示す. 窓打ち状況下で, ワイブル分布を仮定した場合の結果は, 図 5.8 – 5.11 に示す. μ_0 が大きい場合や, サンプルサイズが小さい場合, 阿部・鎌倉 (2016) の方法によって, Rootzén & Zholud 法と比べ, 標準誤差が改善している. この傾向は特に尺度パラメータ η の推定に関して顕著である. Rootzén & Zholud タイプの打ち切り (図 5.4 – 5.7), および窓打ちりの状況下 (図 5.8 – 5.11) 両者において, 同様の傾向が見られた.

Rootzén & Zholud タイプの打ち切りで, ガンマ分布を仮定した場合については図 5.12 – 5.15 に示す. 窓打ち状況下で, ガンマ分布を仮定した場合の結果は, 図 5.16 – 5.19 に示す. μ_0 が大きい場合や, サンプルサイズが小さい場合, 阿部・鎌倉 (2016) の方法によって, Rootzén & Zholud 法と比べ, 標準誤差が改善している. Rootzén & Zholud タイプの打ち切り (図 5.12 – 5.15), および窓打ちりの状況下 (図 5.16 – 5.19) 両者において, 同様の傾向が見られた.

表 5.1 Rootzén & Zholud 法と阿部・鎌倉 (2016) の方法の分散の比較.

	n	分散の比率
$\mu_0 = 0.2$	25	1.03
	50	1.03
	100	1.03
	250	1.03
	1000	1.03
	∞	1.03
	$\mu_0 = 1$	25
50		1.13
100		1.13
250		1.13
1000		1.13
∞		1.12
$\mu_0 = 5$		25
	50	1.39
	100	1.35
	250	1.33
	1000	1.32
	∞	1.31

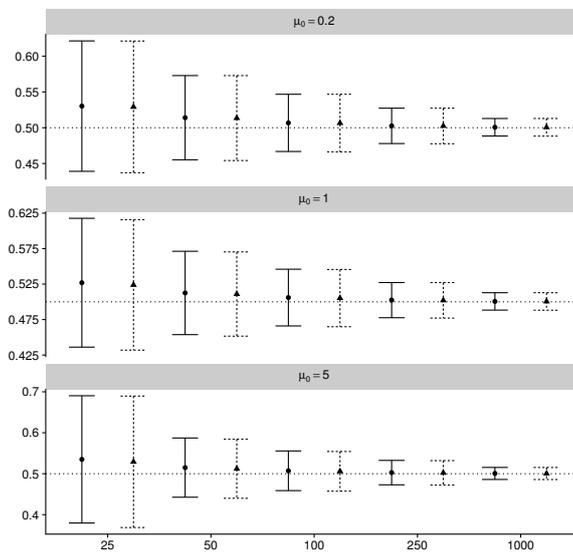


図 5.8 形状パラメータ m の推定値, ワイブル分布 ($m = 0.5$) を仮定し, 窓打ちの場合

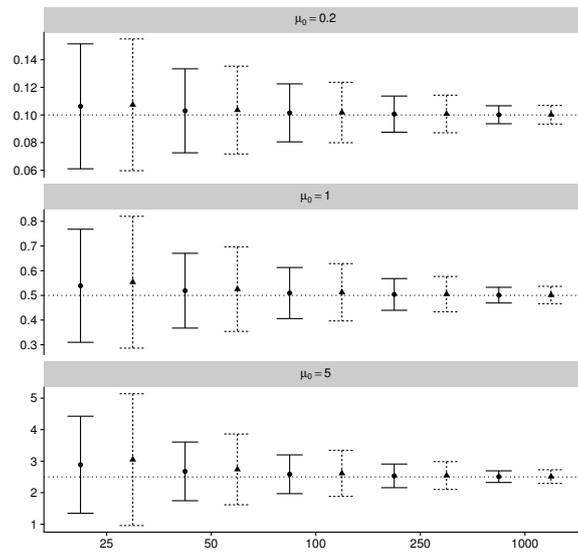


図 5.9 尺度パラメータ η の推定値. ワイブル分布 ($m = 0.5$) を仮定し, 窓打ちの場合.

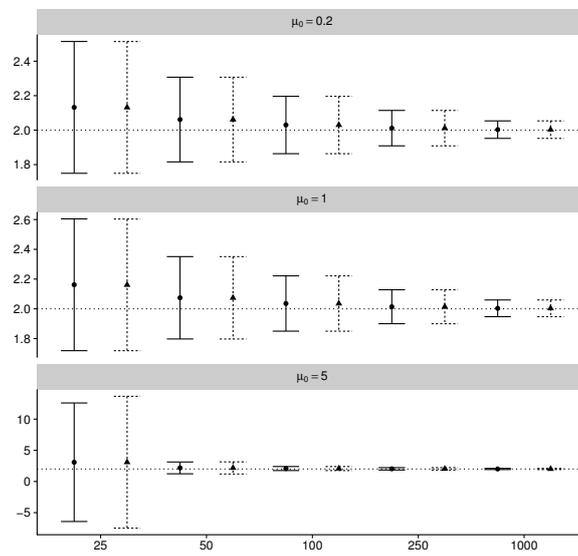


図 5.10 形状パラメータ m の推定値. ワイブル分布 ($m = 2$) を仮定し, 窓打ちの場合.

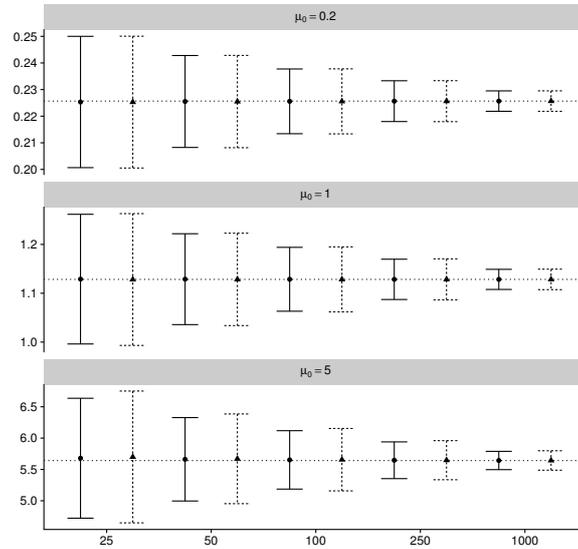


図 5.11 尺度パラメータ η の推定値. ワイブル分布 ($m = 2$) を仮定し, 窓打ち切りの場合.

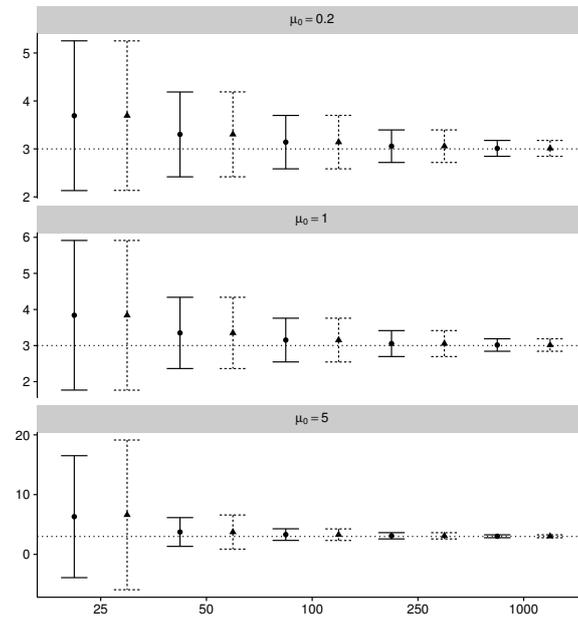


図 5.12 形状パラメータ k の推定値. ガンマ分布 ($k = 3$) を仮定し, Rootzén & Zhould タイプの打ち切りの場合.

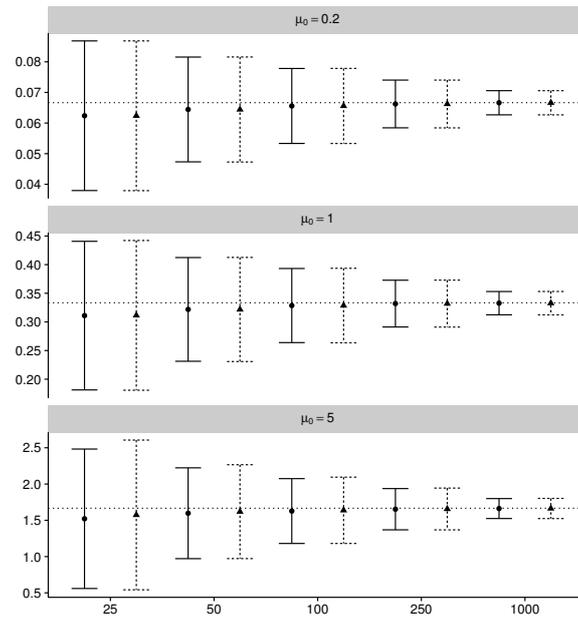


図 5.13 尺度パラメータ σ の推定値. ガンマ分布 ($k = 3$) を仮定し, Rootzén & Zhould タイプの打ちりの場合.

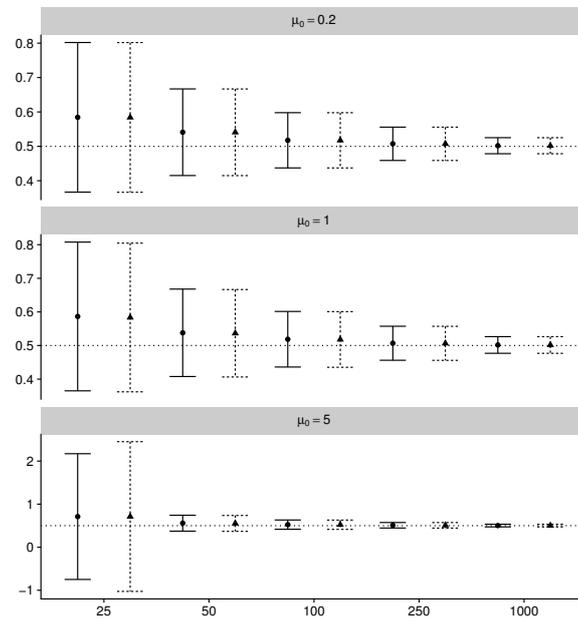


図 5.14 形状パラメータ k の推定値. ガンマ分布 ($k = 0.5$) を仮定し, Rootzén & Zhould タイプの打ちりの場合.

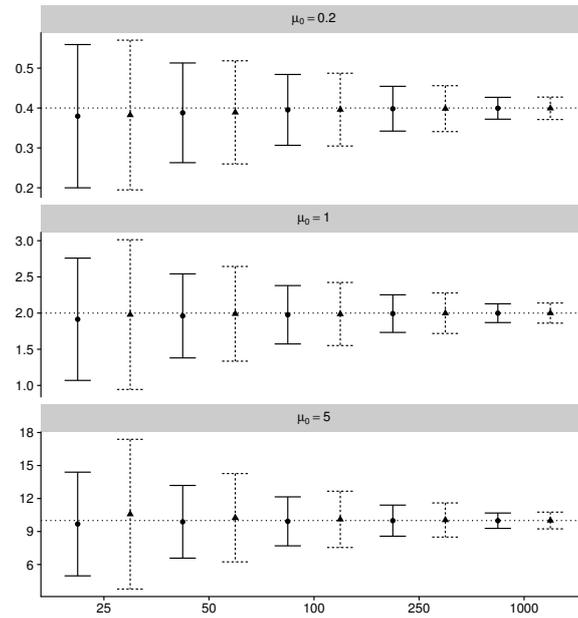


図 5.15 尺度パラメータ σ の推定値. ガンマ分布 ($k = 0.5$) を仮定し, Rootzén & Zhould タイプの打切りの場合.

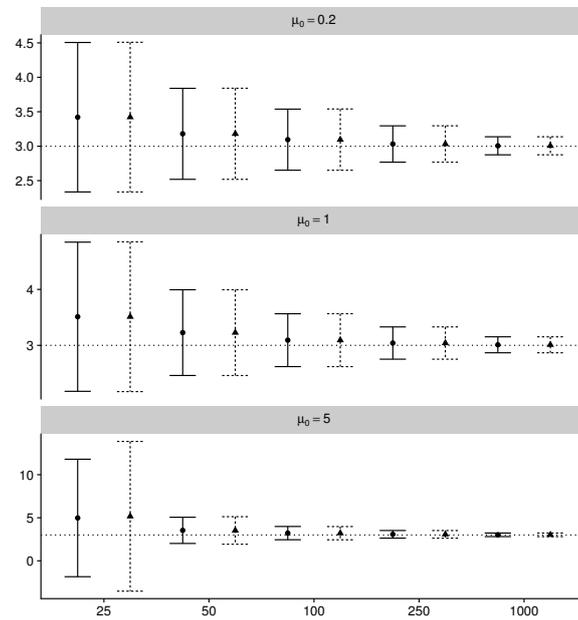


図 5.16 形状パラメータ k の推定値. ガンマ分布 ($k = 3$) を仮定し, 窓打切りの場合.

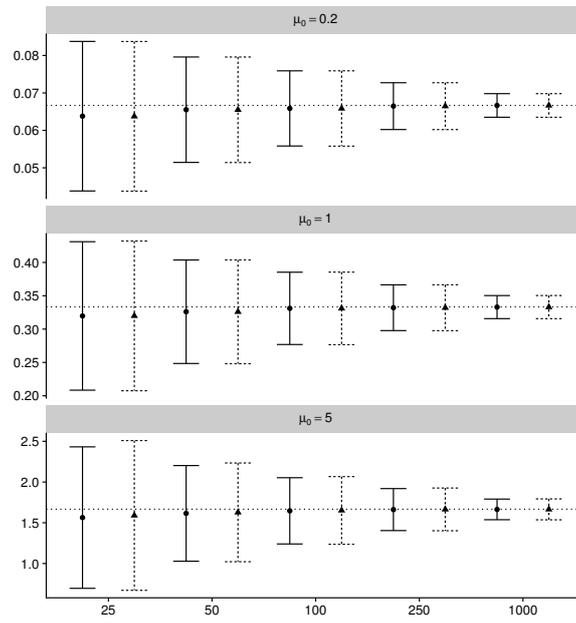


図 5.17 尺度パラメータ σ の推定値. ガンマ分布 ($k = 3$) を仮定し, 窓打切りの場合.

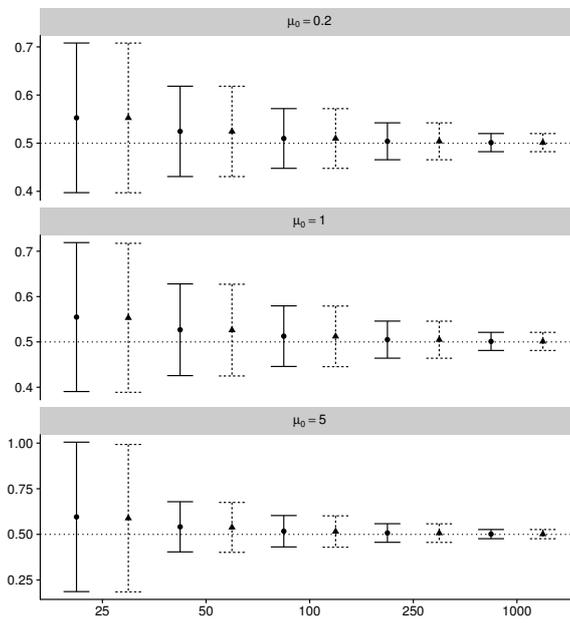


図 5.18 形状パラメータ k の推定値. ガンマ分布 ($k = 0.5$) を仮定し, 窓打切りの打切りの場合.

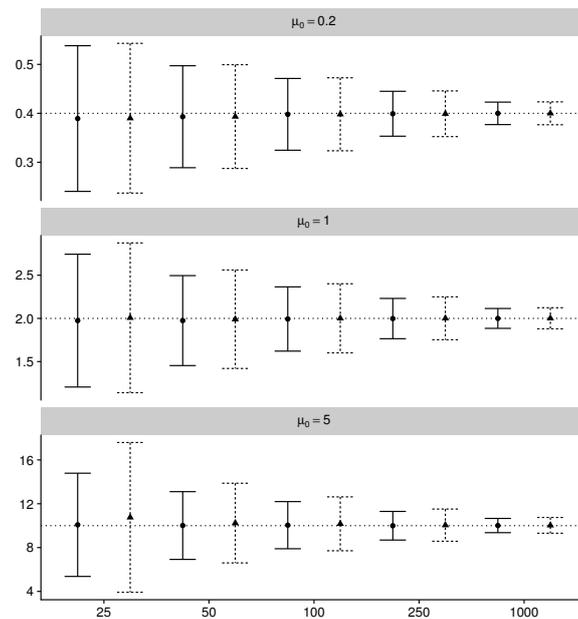


図 5.19 尺度パラメータ σ の推定値. ガンマ分布 ($k = 0.5$) を仮定し, 窓打切りの打切りの場合.

5.5.2 MTTR の推定

ここで再び 2.6.2 節で述べた信頼性工学の話題に立ち返る. Rootzén & Zhoould (2016) は状態 0 の長さの分布のパラメータを推定することに関心があったが, 状態 0 を休止, 状態 1 を稼働とみなせば, 状態 0 の長さの分布の平均は MTTR と呼ばれる. MTBF の推定に関してもまったく同じ方法が適用できる. 本節では 5.5.1 節と同じ窓打切りのシミュレーションを用いて, 分布の平均を求めた場合, どのような結果が得られるか検討する. 推定された状態 0 の分布の平均 (μ_0) の MSE を図 5.20–5.29 に示す. これらの図において, "full" は完全な尤度関数を用いた阿部・鎌倉 (2016) の推定方法, "Rootzen" は Rootzén & Zhoould (2016) の用いた条件付き最尤法に基づく推定方法を指す. MSE で比較すると, サンプルサイズ (観測する窓の数) が小さいときは, 総じて阿部・鎌倉 (2016) の方法が良好な性質を持っている. また, 図 5.22, 5.27 に示した通りワイブル分布の形状パラメータが 0.5 でサンプルサイズが小さいとき, 状態 0 の長さの分布の平均が長いとき, Rootzén & Zhoould (2016) の用いた方法は, 最適化プロセスの収束は確認されたものの極端な値を取ることが多く, 推定そのものが困難である. 図のみやすさのため, サンプルサイズが 50 以上のときのみを抜き出したものを付録に記載する. サンプルサイズが 50 以上の場合についても, 総じて阿部・鎌倉 (2016) の方法が良好な性質を持つ.

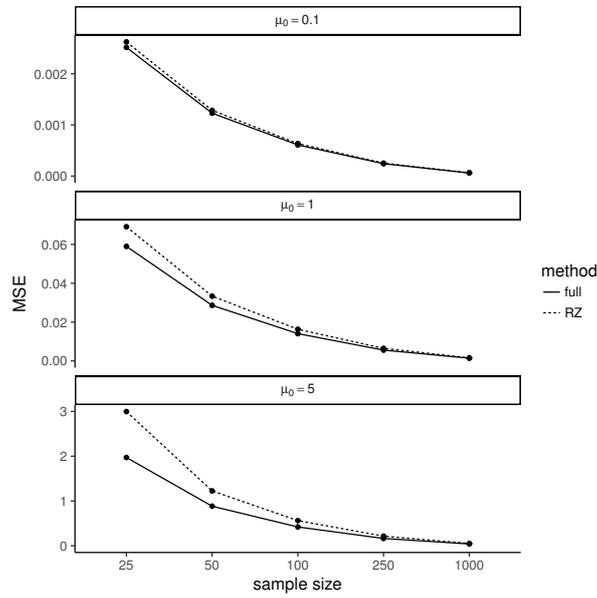


図 5.20 MTTR の推定値の比較. 指数分布を仮定し, Rootzén & Zould タイプの打ちりの場合.

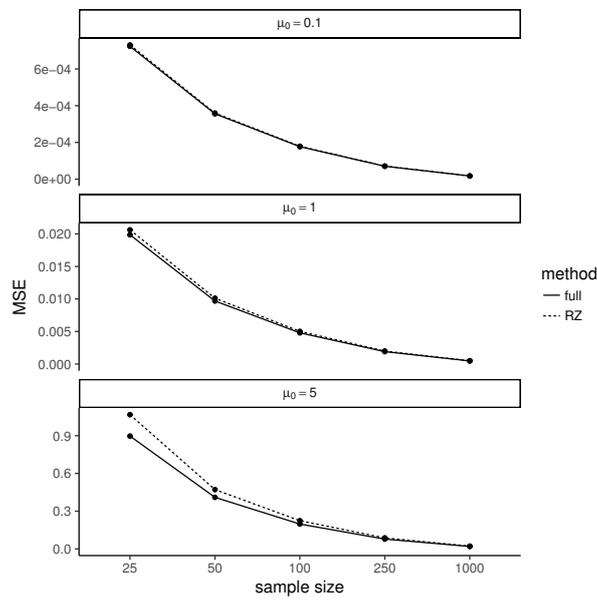


図 5.21 MTTR の推定値の比較. ワイブル分布 ($m = 2$) を仮定し, Rootzén & Zould タイプの打ちりの場合.

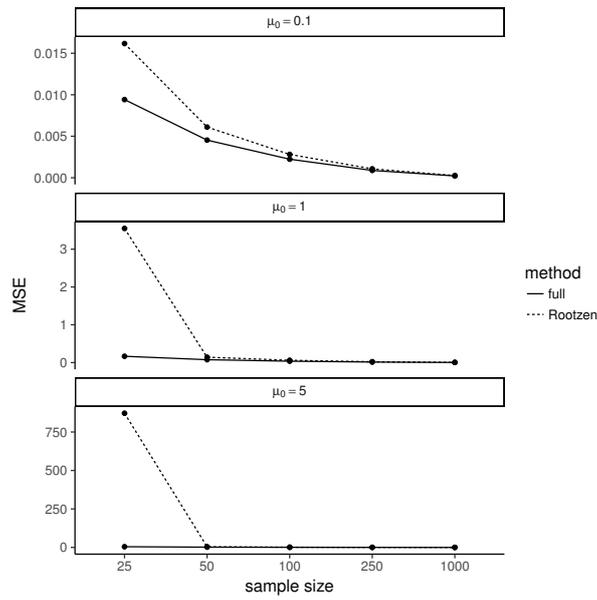


図 5.22 MTTR の推定値の比較. ワイブル分布 ($m = 0.5$) を仮定し, Rootzén & Zhould タイプの打ち切りの場合

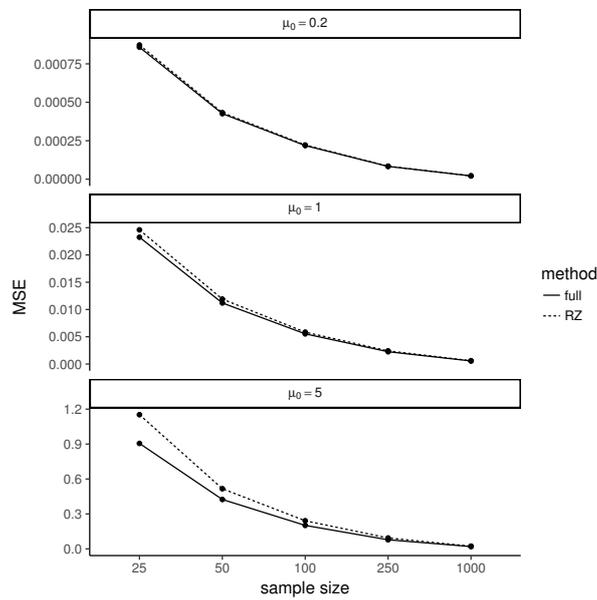


図 5.23 MTTR の推定値の比較. ガンマ分布 ($k = 3$) を仮定し, Rootzén & Zhould タイプの打ち切りの場合.

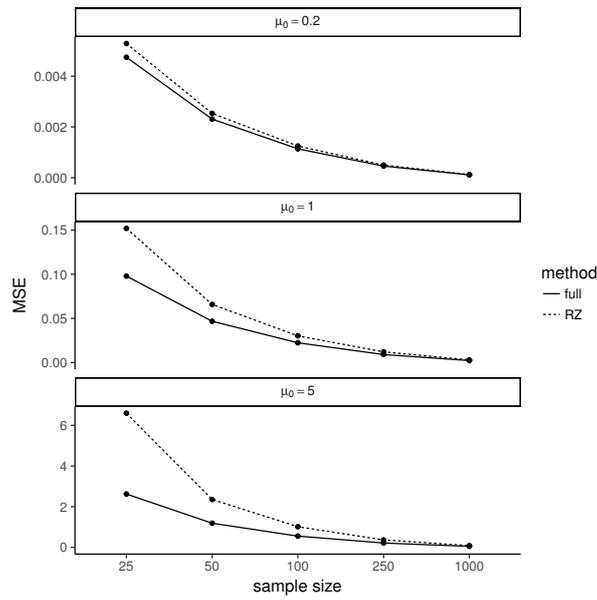


図 5.24 MTTR の推定値の比較. ガンマ分布 ($k = 0.5$) を仮定し, Rootzén & Zhould タイプの打切りの場合,

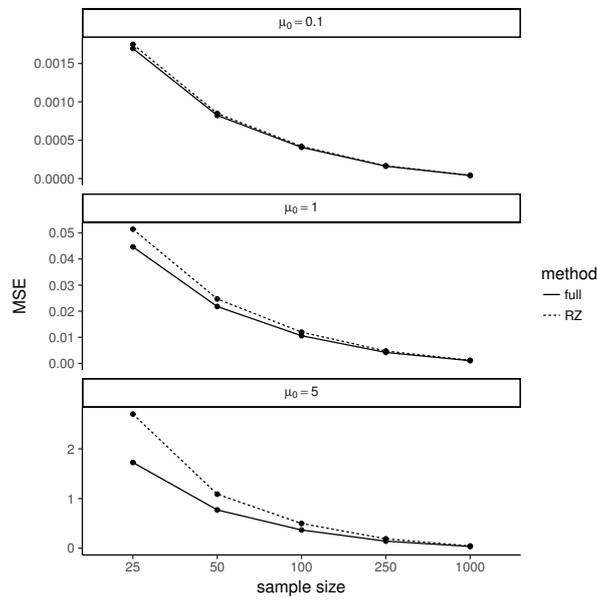


図 5.25 MTTR の推定値の比較. 指数分布を仮定し, 窓打切りの場合

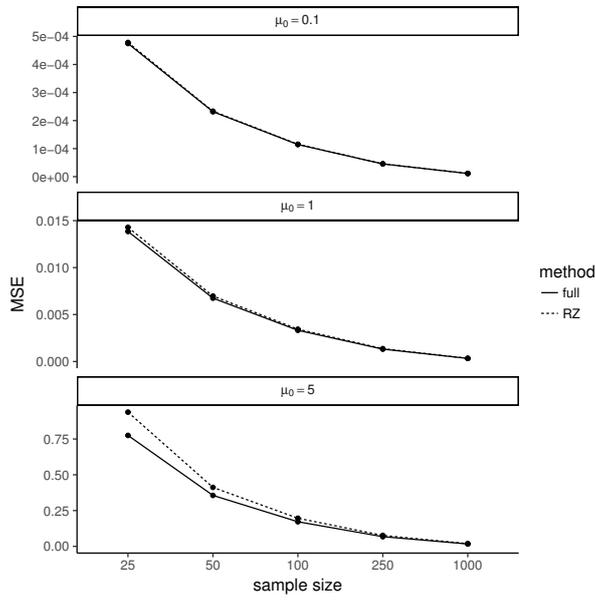


図 5.26 MTTR の推定値の比較. ワイブル分布 ($m = 2$) を仮定し, 窓打ち切りの場合.

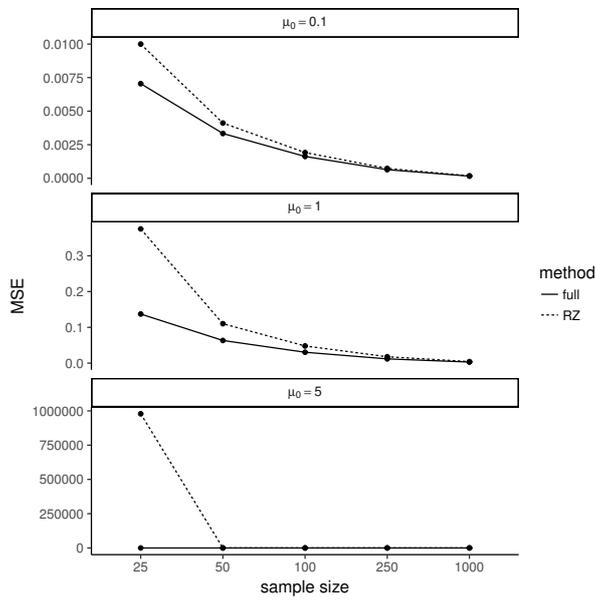


図 5.27 MTTR の推定値の比較. ワイブル分布 ($m = 0.5$) を仮定し, 窓打ち切りの場合.

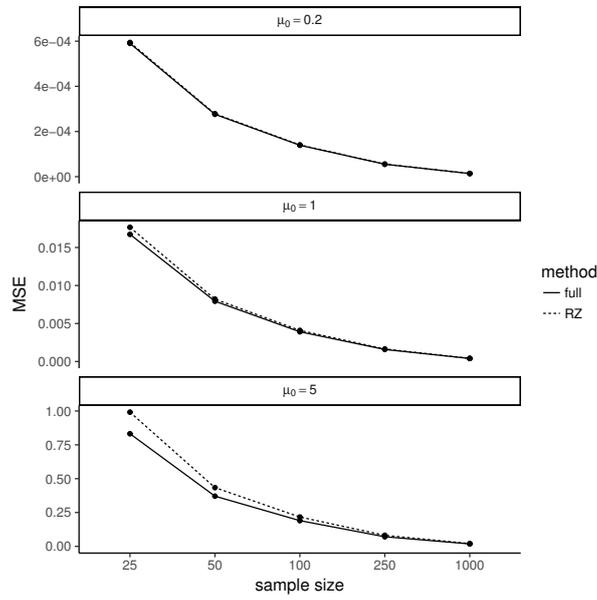


図 5.28 MTTR の推定値の比較. ガンマ分布 ($k = 3$) を仮定し, 窓打切りの場合.

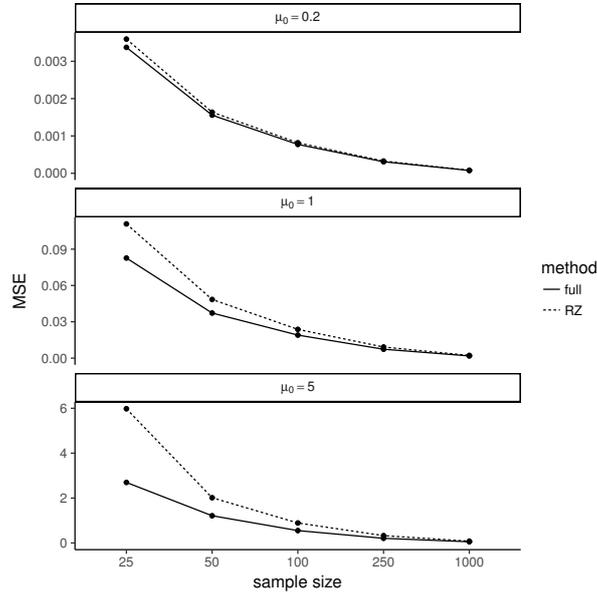


図 5.29 MTTR の推定値の比較. ガンマ分布 ($k = 0.5$) を仮定し, 窓打切りの場合.

5.6 事例研究

本研究の動機は顕微鏡写真を通して観測されたヒートシールの接着面の分析から得られた。研究対象となる製品は、樹脂の素材を熱で溶かして接着するものである。我々は断面の顕微鏡写真から、樹脂の界面がどの程度接着されているかを調査する。

目的は界面の溶着時の温度にともなう変化を明らかにすることである。観測されるのは、界面の長さ、その界面が溶着しているか否か、そして溶着時の温度である。界面は線分と見ることができ、その状態は溶着しているか、していないかの二通りに明確に区別できる。水準は溶着時の温度により、100, 105, 110, 115, 120, 125, 130, 135, 145, 155 (°C) の10段階にわかれ、それぞれ各10枚の顕微鏡写真がある。温度の影響を解析できるモデルでなければならぬため、温度を説明変数にしたパラメトリックモデルを構築する。

特に関心があるのは、ある温度でシールしたとき、界面がどの程度の割合で溶着するかということである。状態1を溶着面、状態0を非溶着面に対応させ、溶着面の長さの平均を全体の長さの平均で割った $\mu_1/(\mu_0 + \mu_1)$ を溶着率と呼ぶことにする。

解析にあたって、引き続き指数分布とワイブル分布によるモデル化を行う。指数分布を仮定した場合、溶着面の長さにモデル式 $F_1(z) = 1 - \exp(-z/b_1)$, $b_1 = \exp(\beta_{1,0} + \beta_{1,1}x)$, 非溶着面の長さにモデル式 $F_0(z) = 1 - \exp(-z/b_0)$, $b_0 = \exp(\beta_{0,0} + \beta_{0,1}x)$ をそれぞれ当てはめ、パラメータ $\beta_{1,0}$, $\beta_{1,1}$, $\beta_{0,0}$, $\beta_{0,1}$ を推定した。ここで x は温度である。

ワイブル分布についても同様に、溶着面の長さにモデル式 $F_1(z) = 1 - \exp(-(z/b_1)^{m_1})$, $b_1 = \exp(\beta_{1,0} + \beta_{1,1}x)$, 非溶着面の長さに $F_0(z) = 1 - \exp(-(z/b_0)^{m_0})$, $b_0 = \exp(\beta_{0,0} + \beta_{0,1}x)$ をそれぞれ当てはめ、パラメータ $m_1, \beta_{1,0}, \beta_{1,1}, m_0, \beta_{0,0}, \beta_{0,1}$ を阿部・鎌倉 (2016) の方法により推定した。この解析の場合、 $\mu_1/(\mu_0 + \mu_1)$ は温度 x の関数であるため、これを $\rho(x)$ と置く。溶着率が温度に応じて変化することを考慮して、(4.2) 式の尤度関数を改めて書くと、以下ようになる。

$$\begin{aligned}
 L \propto & \prod_{i=1}^m \rho(x_i)^{(1-a_i)s_i} \{1 - \rho(x_i)\}^{(1-a_i)(1-s_i)} \\
 & \times \left(\left[\{f_1(\tau_i)\}^{d_i} \{1 - F_1(\tau_i)\}^{1-d_i} \right]^{a_i} \right. \\
 & \times \left. \left[\{g_1(\tau_i)\}^{d_i} \{G_1(\tau_i)\}^{1-d_i} \right]^{1-a_i} \right)^{s_i} \\
 & \times \left(\left[\{f_0(\tau_i)\}^{d_i} \{1 - F_0(\tau_i)\}^{1-d_i} \right]^{a_i} \right. \\
 & \times \left. \left[\{g_0(\tau_i)\}^{d_i} \{G_0(\tau_i)\}^{1-d_i} \right]^{1-a_i} \right)^{1-s_i}. \tag{5.11}
 \end{aligned}$$

各パラメータの推定値は表 5.2 に示す。

溶着率が同じでも、細かい未接着面が多数ある場合と、長い未接着が少数ある場合では、強度に与える影響は異なるため、界面の長さの分布の情報は重要である。指数分布のモデルに比べ、ワイブル分布

表 5.2 指数分布とワイブル分布のモデルにおける各パラメータの推定値と AIC.

	$\beta_{1,0}$	$\beta_{1,1}$	m_1	$\beta_{0,0}$	$\beta_{0,1}$	m_0	AIC
指数分布	-14.92	0.16	-	7.96	-0.06	-	731.18
ワイブル分布	-14.66	0.16	1.31	7.45	-0.05	2.35	678.94

のモデルの AIC が小さくなった. また形状パラメータ m_0 が 1 より十分に大きいことから, 非溶着面の長さは指数分布よりも, 平均の周りに集中した分布であることが分かる.

また, ヒートシールは温度を上げるほど, 接着は容易になるが, 温度を上げすぎると材料の劣化を招くことがあるため, 溶着率が一定に達する最低の温度に関心がある. 温度 x における平均溶着率を $\rho(x)$ で表すと, その推定値 $\tilde{\rho}(x)$ は,

$$\tilde{\rho}(x) = \tilde{\mu}_1 / (\tilde{\mu}_1 + \tilde{\mu}_0)$$

で求められる. $\rho(x)$ の逆関数 $\rho^{-1}(p)$ により, 目標の溶着率 p ($0 < p < 1$) に達する温度を推定することができる. ワイブル分布を仮定した場合,

$$\rho^{-1}(p) = \left\{ \log \left(\frac{1-p}{p} \right) - \left(\log \Gamma \left(1 + \frac{1}{m_0} \right) - \log \Gamma \left(1 + \frac{1}{m_1} \right) \right) - (\beta_{0,0} - \beta_{1,0}) \right\} / (\beta_{0,1} - \beta_{1,1})$$

である. 指数分布を仮定した場合は $\rho^{-1}(p) = \log\{(1-p)/p - (\beta_{0,0} - \beta_{1,0})\} / (\beta_{0,1} - \beta_{1,1})$ である. 例として溶着率 $p = 0.999$ となる温度を求めると, 指数分布のモデルでは 135.61°C , ワイブル分布のモデルでは 136.57°C である.

当てはまりを見るために, (4.4) 式に示したノンパラメトリックな推定量を用いて温度ごとの溶着率を求め, $\rho(x)$ の曲線を重ねてプロットした (図 5.30). 図 5.30 は縦軸に溶着率, 横軸に温度を取り, 指数分布を仮定した場合の平均溶着率を実線で, ワイブル分布を仮定した場合の平均溶着率を点線で示したものである.

5.7 この章のまとめ

本章では, Rootzén & Zholud (2016) による条件付き最尤法を基にした推定 (Rootzén & Zholud 法) と, 阿部・鎌倉 (2016) による最尤法を基にした推定の 2 種類を比較した. 状態 0 の長さの分布のみに関心があり, 状態 1 と状態 0 の長さの分布が独立であっても, 阿部・鎌倉 (2016) の方法が標準誤差で評価して良い性質を持つことがわかった. Rootzén & Zholud 法では状態 1 の長さの比率 ρ に関する情報を使うことができないためである. 特に, 状態 0 の長さの分布と状態 1 の長さの分布がともに指数分布の場合, 漸近相対効率を導くことができる.

事例研究では再生間隔の分布に対して, 状態 1 の長さの分布と状態 0 の長さの分布の尺度パラメータに, それぞれ $\exp(\beta_{1,0} + \beta_1 x)$, $\exp(\beta_{0,0} + \beta_{0,1} x)$ という形で説明変数 x を入れることで, 回帰型のモデルとして利用できることを示した.

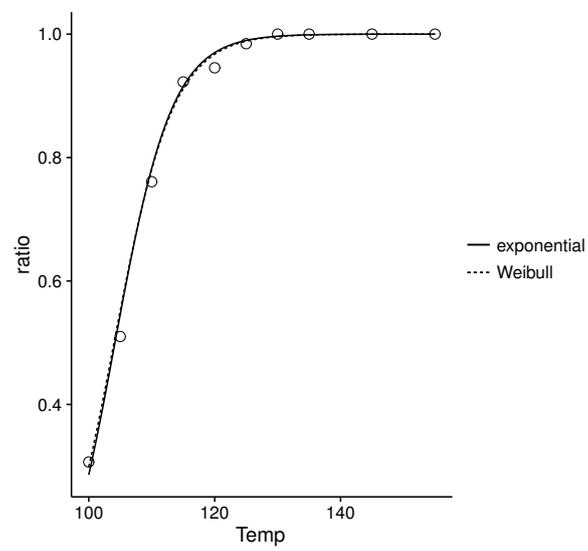


図 5.30 温度と溶着率の関係. 実線：指数分布, 点線：ワイブル分布.

第6章

まとめ

本研究では窓打切りされた確率過程について考察を行った。この章では本研究の成果と応用の可能性、今後の課題について述べる。

はじめに、本研究の成果を要約する。第一に窓打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、バイアス、標準誤差、MSE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。

事例研究ではヒートシールの界面を交代再生過程として表現する新しいモデルを提案し、分析を行った。回帰型のモデルにより、温度ごとの界面の長さの分布を明らかにし、溶着率が一定に達する温度を推定することができた。この成果は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。

最後に今後の課題について述べる。

本研究ではイベントの生起の間隔が独立であることを仮定した。イベント間に従属性がある場合には、均衡分布が簡単には求まらない。しかし現実的には過去のイベントが次のイベントに影響を与えることは考えられるため、より柔軟なモデルの開発が必要である。

次いで、モデルのフィッティングを視覚的に把握する方法についても検討する必要がある。Rootzén& Zholud (2016) では、窓の左端点からはじまる観測と、そうでない観測を分けて Kaplan-Meier 推定量で分布を推定し、パラメトリックなモデルの適合を確認しているが、本研究ではサンプルサイズが小さいため、その手法を適用することはしなかった。

ヒートシールの分析としては、シール強度のデータと関連付けたモデルを考える必要がある。シール強度の分布と界面の長さの分布間にどのような対応関係があるか、より詳細な調査が求められる。

また、本研究はヒートシールの界面の長さの分布を推定することを目的としたため、一次元の空間を対象とした。より多様な空間上のデータを分析するにあたっては2次元への拡張が必要とされる。2次元空間内の線分のデータの分析は Rohrbaugh *et al.* (2002) に例があるが、線分の長さや数の平均

を求めるに留まり, 分布の推定は行われていない. 今後, 2次元空間内の線分を分析するにあたっては, 線分の数, 長さ, 角度の同時分布を推定するという困難な問題を解決しなければならない.

参考文献

- [1] 阿部興・鎌倉稔成 (2014). ランダムなエントリー時間を持つ生存関数の推定. 日本計算機統計学会 第 28 回大会. 論文集. 23–26
- [2] 阿部興・鎌倉稔成 (2016). 窓打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.
- [3] 阿部興・鎌倉稔成 (2017). 窓打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).
- [4] Andersen, P. K., Borgan, O., Gill, R. D. & Keiding, N., (1995) *Statistical Models Based on Counting Processes*. Springer Series in Statistics
- [5] Barlow, R.E. & Preschan, F. (1965). *Mathematical Theory of Reliability*. John Wiley.
- [6] Baxter, A. & Li, L.(1994). Non-parametric confidence intervals for the renewal function and the point availability. *Scandinavian Journal of Statistics.*, Vol. **21**, No.3, 277–287.
- [7] Beichelt, F. E. & Paul Fattil, L. (2002). *Stochastic Processes and Their Applications*. Taylor & Francis.
- [8] Cox, D.R. & Isham, V. (1980). *Point Processes*. Chapman & Hall.
- [9] Cox, D.R. & Miller, H. D. (1965). *The Theory of Stochastic Processes*. Methuen & co ltd.
- [10] Crow, L. H. (1988) Reliability growth estimation with missing data - II. *Annual RELLIABILITY AND MAINTAINABILITY Symposium*.
- [11] Feller, W. (1971). *Introduction to the Theory of Probability and its Applications, Vol. 2. II (2. Ed.)* New York: Wiley.
- [12] Kulkarni, V.G. (1995). *Modeling and analysis of stochastic system*. Chapman & Hall.
- [13] Laslett, G. M. (1982). The survival curve under monotone density constraints with application to two-dimensional line segment processes. *Biometrika.*, **69**: 153-160.
- [14] Mauldon, M. (1998). Estimating mean fracture trace length and density from observations in convex windows. *Rock Mechanics and Rock Engineering.*, 31(4), 201–216.
- [15] Nelder, J. A. & Mead, R. (1965) A simplex method for function minimization. *Computer Journal.* **7**: 308–313.
- [16] 尾崎俊治 (1996). 確率モデル入門. 朝倉書店.
- [17] Peuter C. D. (2013). Modeling basketball as alternating renewal-reward process and predicting match outcomes. *A thesis submitted to the Department of Mathematics for honors*. Duke

- University.
- [18] Pham-Gia, T. & Turkkan(1999). System availability in a gamma alternating renewal process. *Naval Reserch Logistics.*, Vol. **46**. 822–846.
- [19] Rohrbaugh Jr, M. B., Dunne, W. M., & Mauldon, M. (2002). Estimating fracture trace intensity, density, and mean length using circular scan lines and windows. *AAPG bulletin.*, 86(12), 2089–2104.
- [20] Rootzén H. & Zholud, D. (2016). Tail Estimation for Window-Censored Processes. *Technometrics.*, **58**:1, 95–103.
- [21] Ross, S. M. (1995) *Stochastic Processes*. Wiley; 2nd edition.
- [22] Turnbull, B. W. (1976) *The empirical distribution function with arbitrarily grouped, censored and truncated data*. J. R. Statist. Soc. B 38, 290–295.
- [23] Vardi, Y. (1982). Nonparametric estimation in renewal processes. *The Annals of statistics.*, Vol **10**, 772–785.
- [24] Van Zwet, E. W. (2004) Laslett’s line segment problem. *Bernoulli.*, Volume **10**, Number **3**, 377–396.
- [25] Zhao, Y. & Nagaraja, H. N. (2011). Fisher information in window censored renewal process data and its applications. *Annals of the Institute of Statistical Mathematics*, **63**:4, 791-825.
- [26] Zhu, Y., Yashchin E. & Hosking, J. R. M. (2014) Parametric estimation for window censored recurrence data. *Technometrics.*, **56**:1, 55–66.

付録

A シミュレーションによる Laslett (1982) の推定量の検証

Laslett (1982) の推定量について、3.6 節で行ったシミュレーションでは Mauldon(1998) との比較のため、平均の推定値についての RMSE のみを記載した。ここでは各パラメータの平均と標準誤差を記載する。

ガンマ分布に関しては、形状パラメータ $k = 2$ のときの推定が、 $k = 0.5$ のときに比べて難しいことが分かる。ワイブル分布に関しても、形状パラメータ $m = 2$ のときの推定が、 $m = 0.5$ のときに比べて難しいことが分かる。

表 A.1 ガンマ分布のパラメータの推定 ($k = 2$ のとき)

真値				平均		標準誤差	
k	σ	λ	μ	k	σ	k	σ
2.00	0.05	50.00	0.10	2.11	0.05	0.43	0.01
2.00	0.25	50.00	0.50	2.12	0.25	0.46	0.06
2.00	0.50	50.00	1.00	2.16	0.49	0.52	0.12
2.00	1.00	50.00	2.00	2.31	0.97	1.51	0.29
2.00	0.05	100.00	0.10	2.06	0.05	0.28	0.01
2.00	0.25	100.00	0.50	2.06	0.25	0.31	0.04
2.00	0.50	100.00	1.00	2.07	0.50	0.33	0.09
2.00	1.00	100.00	2.00	2.10	0.99	0.40	0.19
2.00	0.05	200.00	0.10	2.03	0.05	0.19	0.01
2.00	0.25	200.00	0.50	2.03	0.25	0.21	0.03
2.00	0.50	200.00	1.00	2.03	0.50	0.23	0.06
2.00	1.00	200.00	2.00	2.04	1.00	0.26	0.13

表 A.2 ガンマ分布のパラメータの推定 ($k = 0.5$ のとき)

真値				平均		標準誤差	
k	σ	λ	μ	k	σ	k	σ
0.50	0.20	50.00	0.10	0.52	0.20	0.09	0.05
0.50	1.00	50.00	0.50	0.52	1.00	0.09	0.22
0.50	2.00	50.00	1.00	0.52	2.01	0.09	0.43
0.50	4.00	50.00	2.00	0.52	3.99	0.09	0.82
0.50	0.20	100.00	0.10	0.51	0.20	0.06	0.03
0.50	1.00	100.00	0.50	0.51	1.00	0.06	0.16
0.50	2.00	100.00	1.00	0.51	2.00	0.06	0.29
0.50	4.00	100.00	2.00	0.51	4.00	0.06	0.58
0.50	0.20	200.00	0.10	0.51	0.20	0.04	0.02
0.50	1.00	200.00	0.50	0.50	1.00	0.04	0.11
0.50	2.00	200.00	1.00	0.50	2.00	0.04	0.21
0.50	4.00	200.00	2.00	0.50	4.00	0.04	0.40

表 A.3 ワイブル分布のパラメータの推定 ($m = 2$ のとき)

真値				平均		標準誤差	
m	η	λ	μ	m	η	m	η
2.00	0.11	50.00	0.10	2.06	0.11	0.25	0.01
2.00	0.56	50.00	0.50	2.76	2.25	3.04	0.22
2.00	1.13	50.00	1.00	2.10	1.13	0.38	0.10
2.00	2.26	50.00	2.00	2.76	2.25	3.04	0.22
2.00	0.11	100.00	0.10	2.03	0.11	0.17	0.01
2.00	0.56	100.00	0.50	2.12	2.26	0.70	0.15
2.00	1.13	100.00	1.00	2.04	1.13	0.24	0.07
2.00	2.26	100.00	2.00	2.12	2.26	0.70	0.15
2.00	0.11	200.00	0.10	2.01	0.11	0.12	0.00
2.00	0.56	200.00	0.50	2.04	2.26	0.22	0.11
2.00	1.13	200.00	1.00	2.03	1.13	0.17	0.05
2.00	2.26	200.00	2.00	2.04	2.26	0.22	0.11

表 A.4 ワイブル分布のパラメータの推定 ($m = 0.5$ のとき)

真値				平均		標準誤差	
m	η	λ	μ	m	η	m	η
0.50	0.05	50.00	0.10	0.51	0.05	0.06	0.01
0.50	0.25	50.00	0.50	0.51	0.26	0.05	0.06
0.50	0.50	50.00	1.00	0.51	0.51	0.05	0.12
0.50	1.00	50.00	2.00	0.51	1.02	0.05	0.23
0.50	0.05	100.00	0.10	0.51	0.05	0.04	0.01
0.50	0.25	100.00	0.50	0.50	0.25	0.04	0.04
0.50	0.50	100.00	1.00	0.50	0.51	0.03	0.08
0.50	1.00	100.00	2.00	0.50	1.01	0.04	0.16
0.50	0.05	200.00	0.10	0.50	0.05	0.03	0.01
0.50	0.25	200.00	0.50	0.50	0.25	0.02	0.03
0.50	0.50	200.00	1.00	0.50	0.50	0.02	0.06
0.50	1.00	200.00	2.00	0.50	1.01	0.02	0.11

B MTTR の推定量の比較

5.5.2 節においてシミュレーションした, 状態 0 の長さの分布に形状パラメータ $m = 0.5$, 尺度パラメータ $\eta = 2.5$ のワイブル分布を仮定した場合について, 平均の推定値の MSE を図 B.1–B.2 に示す.

これらの図において, "full" は完全な尤度関数を用いた阿部・鎌倉 (2016) の推定方法, "RZ" は Rootzén & Zould (2016) の用いた条件付き最尤法に基づく推定方法を指す. 図のみやすさのため, サンプルサイズが 50 以上のときのみを抜き出したものを付録に記載する. 総じて阿部・鎌倉 (2016) の方法が良好な性質を持つ.

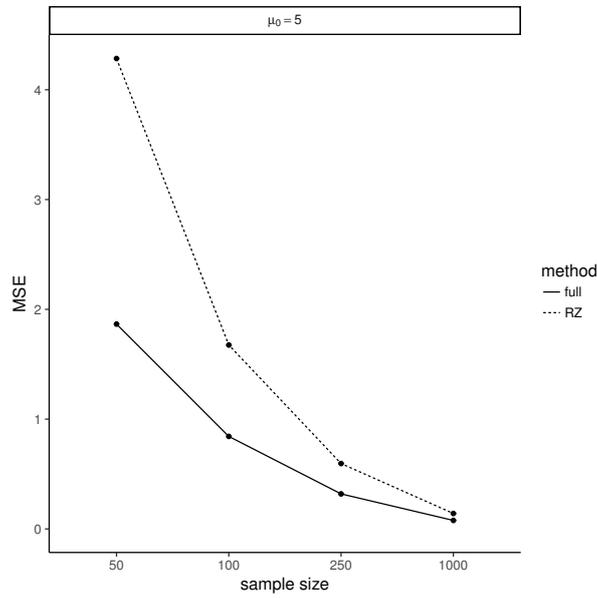


図 B.1 MTTR の推定値の比較. ワイブル分布 ($m = 0.5$) を仮定し, Rootzén & Zhould タイプの打ち切りの場合.

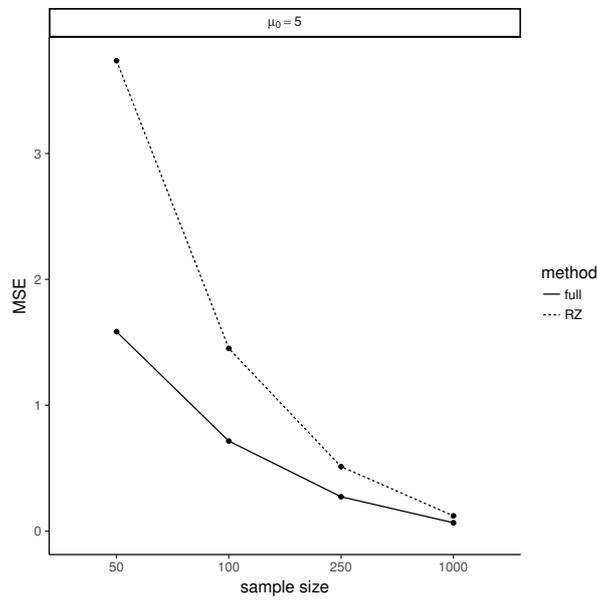


図 B.2 MTTR の推定値の比較. ワイブル分布 ($m = 0.5$) を仮定し, 窓打ち切りの場合.