

# パラメトリック潜在故障時間コンピュータモデルに基づく

## 左側切断・競合リスクデータによるベイズ推定

江村剛志\*

統計数理研究所・医療健康データ科学研究センター

道前洋史

北里大学・薬学部

生存時間解析において、打ち切り・切断・競合リスクといった不完全データの取り扱い、古典的ではあるが重要なトピックである。近年、競合リスクを伴う左側切断データの解析で、Kundu et al. (2017) は独立な競合リスクモデルによる統計学的手法を提案した。彼らの提案手法は、競合するイベント（リスク）が互いに独立であるという仮定の下、左切断を扱うことが出来るが、この独立性の仮定は多くの実証研究で成り立たない。そこで、Michimae & Emura (2022) は独立性の仮定を緩め、相関のある競合リスクのためのコンピュータモデルを提案したが、相関パラメータの推定が不安定であるため最尤法には困難が伴うことが示された。そこで、本研究 (Michimae et al. 2024) では、左切断・競合リスクモデルのパラメータ推定をベイズ推定法で行うことを提案する。シミュレーション実験を行い、ベイズ法が最尤法よりも優れていることを示す。最後に、データセットの解析例を紹介する。

## 1. はじめに

機器が現場で使用されている状態、つまりフィールド試験で収集された故障データ（フィールド寿命データ）に基づく解析は故障時間を測定し、機器の現場における信頼性を解析することにある。これは、すべての機器を一斉に使用開始する実験室試験と異なり、開始時期はランダムである（機器はランダムに設置されると仮定する）。このデータ収集法において、収集期間前に早期故障が発生するため、標本の切り捨て（左切断）が起こる。このような左切断を伴うフィールド寿命データの統計解析法は、実験室試験が困難な長い寿命を持つ機器の信頼性解析に対して提案されてきた (Hong et al. 2009; Emura & Shiu 2016; Emura & Michimae 2022; Balakrishnan & Mitra 2012; Mitra et al. 2021)。特に、Hong et al. (2009) は左切断を考慮した電力変換器のフィールド寿命データをワイブル分布に当てはめる最尤法を導入した。

Kundu et al. (2017)は上述の最尤法を競合リスク下のデータに拡張した。しかしながら、彼らの手法は、リスク間の独立性を仮定した解析法である。この独立性の仮定は多くの実証研究で成立しない。この仮定を緩めたモデルとして Wu et al. (2021)は、マーシャル・オルキン型の2変量レイリー分布モデルを用いて従属競合リスクのモデルへ最尤法を拡張している。同様に、Wang et al. (2022)は一般化逆指数分布を用いた2変量従属リスクモデルへ最尤法を拡張している。しかしながら、これらの従属リスクモデルは、特定の相関構造と特定の周辺分布を仮定した解析であり、信頼性解析において必要な周辺分布であるワイブル・対数正規・ガンマ分布などを含まない特殊なモデルである。

相関構造と周辺分布を柔軟にモデリングするにはコピュラモデルが有用となる。左切断が存在する状況下で、Michimae & Emura (2022)は、相関のある競合リスクのためのコピュラモデルを提案し、最尤法を定式化した。コピュラモデルにおいて、周辺分布は柔軟に選ぶことが出来るため、ワイブル・対数正規・ガンマ分布などが採用された。しかしながら、相関パラメータの推定が不安定であるため最尤法には困難が伴うことが示された。そこで、本研究 (Michimae et al. 2024)では、左切断・競合リスクモデルのパラメータ推定をベイズ推定法で行うことを提案する。特に、事後密度や事前分布の設定を定式化する。シミュレーション実験を行い、提案するベイズ法と最尤法の比較を行う。最後に、データセットの解析例を紹介する。

以下、本講演で用いる競合リスクのための潜在故障時間モデルによるデータの構造 (2 節) と、そのコピュラモデリング法 (3 節) を紹介する。提案手法であるベイズ推定法は4節で説明する。シミュレーションとデータ解析は講演で発表する。

## 2. 左切断・競合リスクのモデル

まず、Cox (1959)により提案された古典的なアプローチである「潜在故障時間」による競合リスクモデルを定義する。このモデルに関する日本語の解説は (江村&道前 2021 ; 道前 2023 ; 武富・山本 2023) があり、主に信頼性工学で広く利用されている。

いま、 $(T_1, T_2)$ を2つの異なる原因による故障時間とする。これらの周辺生存関数を $S_i(t) = P(T_i > t)$ ,  $i \in \{1, 2\}$ とおく。この生存関数は連続であるとし、確率密度関数 $f_i(t) = -dS_i(t)/dt$ ,  $i \in \{1, 2\}$ を持つと仮定する。このように2つの故障時間を考え、その先に起こる故障時間 $T_{12} \equiv \min(T_1, T_2)$ が観測されるモデルを潜在故障時間モデルという。例えば、 $T_1 < T_2$ のとき、観測される故障時間は $T_{12} = T_1$ であり、故障時間 $T_2$ は観測されない。よって、ペア $(T_1, T_2)$ のどちらか一方のみしか観測されない。ペア $(T_1, T_2)$ が同時に観測されないことから、その相関の推定が難しいことがわかる。

図1は左切断・競合リスクのデータ構造を表している。データ収集期間中にランダムに設置される機器を考える（図1の時間軸の全体に設置時間が分布している）。しかし、設置された全ての機器が観測されるわけではなく、早期に設置され、早期に故障した機器は切り捨てられ観測されない。一方、データ収集期間（図1）開始後に設置された機器は必ず観測される。このサンプリングバイアス（早期故障が切り捨てられやすい）を補正するには、打ち切り時間（図1の $c$ ）を考慮した解析が必要となる。すなわち、機器が観測されるためには、条件 $T_{12} \geq \tau$ を満足する必要がある、この条件を考慮した統計的推測が必要となる。

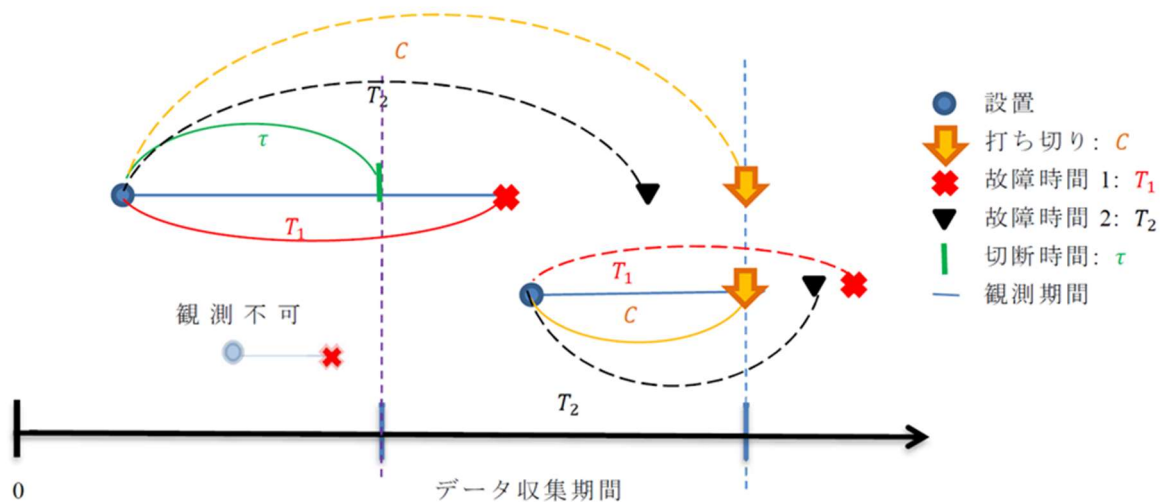


図1；左切断・競合リスクデータ。

上記のデータ収集においては、データ収集期間は限られるため、切断に加え、打ち切りもある。図1にあるよう、打ち切り時間を $C$ で表すとすると、このとき観測される時間は $\min(T_{12}, C)$ と表せる。従って、左切断と競合リスク以外にも、右側打ち切りを考量した統計解析が必要となる。

### 3. 競合リスクのコピュラモデル

Escarela & Carriere(2001)は従属競合リスク解析のために故障時間 $T_1$ と $T_2$ の相関構造に対して、次のようなコピュラモデルを提案した

$$P(T_1 > t_1, T_2 > t_2) = C_\theta(S_1(t_1), S_2(t_2)),$$

ここで $C_\theta$ はパラメータ $\theta$ をもつコピュラ(Nelsen 2006)である。コピュラモデルは競合リスク間の従属構造をモデリングする有用なツールであると同時に、信頼性解析においても有用である(太田 2023)。本講演において紹介するコピュラは、モデルの

複雑性を避けるため、次のクレイトンコピュラに限定する。

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}, \quad \theta > 0, \quad 0 < u, v < 1.$$

ここで $\theta$ は $T_1$ と $T_2$ の相関の強さを定めるパラメータである。ケンドール順位相関に変換すると $\theta/(\theta + 2)$ となる。従って、 $T_1$ と $T_2$ は正相関を表し、 $\theta$ の値が大きいほど相関は強くなる。独立のケースは極限 $\theta \rightarrow 0$ に対応し、独立コピュラ $C(u, v) = uv$ に収束する。この意味において、Kundu et al. (2017)の独立モデルを極限として表現する。

周辺分布は様々なモデルを適用可能であるが、本講演では次のワイブル分布

$$S_i(t) = \exp(-\lambda_i t^{\alpha_i}), \quad i = 1, 2.$$

を考え、ここに $\lambda_i > 0$ は尺度パラメータ（故障の強度を表す）、 $\alpha_i > 0$ はハザードの減少( $\alpha_i < 1$ )または増加( $\alpha_i > 1$ )の形状を表す。

以上のクレイトンコピュラモデルにおける未知パラメータは $(\lambda_1, \alpha_1, \lambda_2, \alpha_2, \theta)$ となる。従属競合リスクと左切断の下で観測されたデータにより構成された尤度関数を $L(\lambda_1, \alpha_1, \lambda_2, \alpha_2, \theta | \text{Data})$ とする。尤度関数の具体的な式と、最尤推定量の定義・計算法はMichimae & Emura (2022)を参照されたい。

#### 4. 提案手法（ベイズ推定法）

2節で説明したデータ構造の下で得られた「左切断・競合リスクデータ」を基に、未知パラメータ $(\lambda_1, \alpha_1, \lambda_2, \alpha_2, \theta)$ を推定することを提案する。Michimae & Emura (2022)では、最尤法を提案したが、相関パラメータの推定が不安定であるため最尤法には困難が伴うことが示された。本研究 (Michimae et al. 2024) では、最尤法に代わり、ベイズ推定法を提案する。ベイズ推定法は、与えられたデータの情報と事前分布の情報を統合した推定法である。

より具体的には、データの情報を基にした尤度に、事前分布の密度を掛け合わせた事後分布を基に推測を行う。すなわち、事後密度は

$$f(\lambda_1, \alpha_1, \lambda_2, \alpha_2, \theta | \text{Data}) \propto L(\lambda_1, \alpha_1, \lambda_2, \alpha_2, \theta | \text{Data}) f(\lambda_1) f(\alpha_1) f(\lambda_2) f(\alpha_2) f(\theta)$$

であり、ここに右辺の $L(\cdot | \text{Data})$ は尤度関数、 $f(\cdot)$ は事前密度である。

尤度を書き下すため、 $n$ 個の標本に基づくデータを $\{(t_i, \tau_i, v_i); i = 1, 2, \dots, n\}$ と記す。ここに $t_i = \min(T_{1i}, T_{2i}, C_i)$ は観測される生存時間であり、 $(T_{1i}, T_{2i})$ は2つの潜在故障時間のペア、 $C_i$ は独立な打ち切り時間である。また、 $\tau_i$ は切断時間である。

ワイブル分布の下での $n$ 個の標本に基づく尤度関数は下のようになる：

$$\begin{aligned}
 L(\alpha_1, \alpha_2, \lambda_1, \lambda_2, \theta | \text{Data}) &= \prod_{i \in l_1} \left\{ \alpha_1 \lambda_1 t_i^{\alpha_1 - 1} u_i^{-\theta} (u_i^{-\theta} + v_i^{-\theta} - 1)^{-(1+1/\theta)} \right\} \\
 &\times \prod_{i \in l_2} \left\{ \alpha_2 \lambda_2 t_i^{\alpha_2 - 1} v_i^{-\theta} (u_i^{-\theta} + v_i^{-\theta} - 1)^{-(1+1/\theta)} \right\} \times \prod_{i \in l_0} (u_i^{-\theta} + v_i^{-\theta} - 1)^{-1/\theta} \\
 &\times \prod_{v_i=0} (x_i^{-\theta} + y_i^{-\theta} - 1)^{1/\theta}.
 \end{aligned}$$

ここに $u_i = \exp(-\lambda_1 t_i^{\alpha_1})$ ,  $v_i = \exp(-\lambda_2 t_i^{\alpha_2})$ ,  $x_i = \exp(-\lambda_1 \tau_i^{\alpha_1})$ ,  $y_i = \exp(-\lambda_2 \tau_i^{\alpha_2})$ である。また、積の下にある $l_i$ は $i(= 1, 2)$ 番目のイベントを経験した集合、 $l_0$ は打ち切られた集合である。コピュラパラメータの事前密度 $f(\theta)$ は $[-1, \infty)$ 上の一様分布とし、その他の事前分布は表1に示す。パラメータの点推定値は事後メディアン、95%信用区間は7.5%と97.5%点を用いる。これら数値計算はStanによるMCMCで行う。

表1：ワイブルパラメータの事前分布とその超パラメータの設定

Prior for Weibull paramaters	Event1	Event2
Gamma distribution	$\lambda_1 \sim G(0.001, 0.001)$	$\lambda_2 \sim G(0.001, 0.001)$
	$\alpha_1 \sim G(0.001, 0.001)$	$\alpha_2 \sim G(0.001, 0.001)$
Inverse-Gamma distribution	$\lambda_1 \sim \text{inv-}G(0.001, 0.001)$	$\lambda_2 \sim \text{inv-}G(0.001, 0.001)$
	$\alpha_1 \sim \text{inv-}G(0.001, 0.001)$	$\alpha_2 \sim \text{inv-}G(0.001, 0.001)$
improper Unifrom distribution	$\lambda_1 \sim U(0, \infty)$	$\lambda_2 \sim U(0, \infty)$
	$\alpha_1 \sim U(0, \infty)$	$\alpha_2 \sim U(0, \infty)$
half Normal distribution	$\lambda_1 \sim N(0, 1000)$	$\lambda_2 \sim N(0, 1000)$
	$\alpha_1 \sim N(0, 1000)$	$\alpha_2 \sim N(0, 1000)$
half Cauchy distribution	$\lambda_1 \sim C(0, 5)$	$\lambda_2 \sim C(0, 5)$
	$\alpha_1 \sim C(0, 5)$	$\alpha_2 \sim C(0, 5)$

## 参考文献

1. Hong Y, et al. (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *Ann Appl Stat* **3**: 857-879.
2. Emura T, Michimae H (2022). Left-truncated and right-censored field failure data: review of parametric analysis for reliability. *Qual Reliab Eng Int*, **38**(7), 3919-3934.
3. Balakrishnan N, Mitra D (2012). Left truncated and right censored Weibull data and likelihood inference with an illustration. *Comp Stat Data Anal* **56**: 4011-4025.

4. 江村剛志, & 道前洋史. (2021). 左切断・右打ち切りのあるフィールド寿命データ解析~ パラメトリックモデルに基づく統計的推論と信頼性の予測~. 電子情報通信学会技術研究報告; 信学技報, **121**(216), 7-12.
5. Emura T, Shiu-SK (2016) Estimation and model selection for left-truncated and right-censored lifetime data with application to electric power transformers analysis, *Commun Stat-Simul* **45**(9): 3171–89.
6. Mitra D, et al. (2021). Likelihood analysis and stochastic EM algorithm for left truncated right censored data and associated model selection from the Lehmann family of life distributions. *Jpn J Stat Data Sci* **4**, 1019–1048.
7. Kundu D, et al. (2017). Analysis of left truncated and right censored competing risks data. *Comp Stat Data Anal* **108**, 12-26.
8. Wu K, et al. (2021). Statistical inference of left truncated and right censored data from Marshall–Olkin bivariate Rayleigh distribution. *Mathematics*, **9**, 2703.
9. Wang L, et al. (2022). Inference of dependent left-truncated and right-censored competing risks data from a general bivariate class of inverse exponentiated distributions. *Statistics*, **56**, 347–374
10. Michimae H and Emura T. (2022). Likelihood inference for copula models based on left-truncated and competing risks data from field studies. *Mathematics*, **10**(13), 2163.
11. Michimae H, et al. (2024). Bayesian parametric estimation based on left-truncated competing risks data under bivariate Clayton copula models, *J of Appl Stat*, doi.org/10.1080/02664763.2024.2315458
12. Cox, D. R. (1959). The analysis of exponentially distributed life-times with two types of failure, *Comp Stat Data Anal: Series B*, **21**, 411–421.
13. 道前洋史. (2023). 競合リスクを伴う左側切断・右側打ち切りデータの解析~ 現状と今後の展望について~. 日本統計学会誌, **52**(2), 203-220.
14. 武富奈菜美, & 山本和嬉. (2023). 生存時間解析・信頼性解析のための統計モデル. 日本統計学会誌, **52**(2), 69-112.
15. Escarela, G., Carriere, J. F. Fitting competing risks with an assumed copula. *Stat Methods Med Res*; **12**(4), 333-349.
16. Nelsen, R. B. *An Introduction to Copulas*, Second Edition, 2006, Springer.
17. 太田修平. (2023). 信頼性データ解析への多変量 Farlie-Gumbel-Morgenstern コピュラの応用. 日本統計学会誌, **52**(2), 177-201.

連絡先： 江村剛志 ([takeshiemura@gmail.com](mailto:takeshiemura@gmail.com))

〒190-8562 東京都立川市緑町 10-3

統計数理研究所