

カーネル法に基づいた修理系システムに対するノンパラメトリック推論に関する一考察

広島大学大学院工学研究科 齋藤 靖洋 (Yasuhiro Saito) 土肥 正 (Tadashi Dohi)
Graduate School of Engineering, Hiroshima University

1 はじめに

確率点過程を用いた故障現象のモデル化に基づくシステムの実データ解析について、従来から数多くの研究が行われてきた。確率点過程は、任意の時刻における故障の発生しやすさを表わす条件付き強度関数をモデル化することによって、様々な故障現象の振る舞いを記述することが出来る。これらの確率点過程は、故障時間分布や修理行動の種類によって分類することが可能である [1]。非修理系システムで故障が発生した場合、システムやコンポーネントは新品に取り換えられる。そのため、取り換え時間がシステム稼働時間全体に対して無視出来るほど小さいとするならば、累積故障件数のふるまいを再生過程 (RP) によって記述することが出来る [4]。RP は、それぞれの故障時間間隔が独立で同一な (i.i.d.) 確率分布 (再生分布) に従うという特徴を持つ。一方、修理系システムで故障が発生した場合、コンポーネントの故障状態を通常状態に戻すために修理行動が行われる。最も単純な例として、故障したコンポーネントを故障直前の状態に復旧する修理行動である小修理を考えた場合、その故障過程は非同次ポアソン過程 (NHPP) で表すことが出来る [1]。

実世界の複雑な修理行動を記述するためには、小修理のような理想的かつ単純な修理行動だけでは不十分である。Brown and Proschan [3] は不完全修理モデルを提案し、取り換え若しくは小修理が一定の確率で施される場合において、システムの累積故障件数が NHPP で表されることを示した。これに対して Kijima [7] は不完全修理モデルの概念を大きく拡張し、Kijima モデルと呼ばれる取り換えと小修理の中間的な修理を表現する一般修理モデルを提案した。このモデルでは、故障したコンポーネントを新品への取り換え直後の状態と故障直前の状態の中間的な状態に復旧する修理行動を表すことが出来る。Kijima モデルの統計的推論に関する重要な結果として、Dorado *et al.* [5] はノンパラメトリック推定量を提案し、その漸近的性質を示した。このような確率点過程のノンパラメトリック推定手法は、パラメトリックな関数が事前に特定出来ないような不確実性の下で、修理系システムの故障発生振る舞いを推論するために役立つ。

これに対して、一般的な NHPP と RP の要素を併せ持つ確率点過程であるトレンド再生過程 (TRP) もまた、ある種の一般修理モデルを表現することが可能である [8]。トレンド再生過程は、NHPP の強度関数に対応するトレンド関数と呼ばれる時間変調関数と、RP の要素である再生分布によって特徴付けられる。Heggland and Lindqvist [6] は、RP の要素としてワイブル再生分布を仮定した TRP (W-TRP) に対して、トレンド関数が未知であるが単調な関数であると仮定出来る場合に対して制約付きノンパラメトリック最尤推定量を提案した。一方、Saito and Dohi [10] は、RP の要素としてガンマ再生分布を仮定した TRP (G-TRP) に対して、トレンド関数の制約付きノンパラメトリック最尤推定量についての考察を行った。これとは逆に、Saito and Dohi [11] はトレンド関数は既知であるが再生分布が未知であるという仮定の下、再生分布に対応する故障率関数に対する制約付きノンパラメトリック最尤推定量を提案した。近年、Lindqvist [9] は最尤法と異なるノンパラメトリック推定手法として、W-TRP でトレンド関数が未知である場合に対してカーネル関数に基づく推定アルゴリズムを提案した。根本的なアイディアは、ワイブル再生分布の初期パラメータを与えた上で、カーネル関数を特徴付けるバンド幅を推定し、以降バンド幅を固定して再生分布のパラメータを更新していくことにある。しかしながら提案されたアルゴリズムでは、更新する再生分布のパラメータが必ずしも収束しないという問題が含まれている。これに対して本稿では、トレンド関数は既知であるが再生分布が未知であるという仮定の下で、再生分布に対応する再生密度関数に対してのカーネル

法を適用する。具体的に、いくつかのバンド幅推定手法を用いて再生分布を推定し、シミュレーション実験を通じて提案するカーネル推定アルゴリズムの有効性を評価すると共に、実データ解析への適用例を示す。

2 トレンド再生過程

時刻 $t = 0$ から稼働中の修理系システムに対し、 n 個の故障時刻データを表す確率変数を T_i ($T_1 < T_2 < \dots < T_n$) とおく。 $i-1$ 番目と i 番目の故障時間間隔を表す確率変数を X_i ($i = 1, 2, \dots, n$) とすると、 $X_i = T_i - T_{i-1}$ が成り立つ。ここで、 $T_0 = 0$ かつ $T_1 = X_1$ とする。それらの実現値を t_i ($t_1 < t_2 < \dots < t_n$) 及び x_i ($i = 1, 2, \dots, n$) で表す。故障が発生するたびに修理行動が行なわれ、システムは故障状態から正常状態へ復旧されるが、各修理にかかる時間はシステムの稼働時間と比べて無視出来るほど十分小さいとする。また一般性を失うことなく $n \geq 2$ であり、システムの稼働中に少なくとも 2 回以上の故障が生じるものと仮定する。確率点過程 $\{N(t), t \geq 0\}$ は時刻 t までに発生した累積故障件数を表すとし、 $\lambda(t)$ 及び $\Lambda(t) = \int_0^t \lambda(u) du$ をそれぞれ確率過程の強度関数及び累積強度関数と定義する。ここで、累積強度関数によって時間スケール変換された時刻列 $\Lambda(T_1), \Lambda(T_2), \dots, \Lambda(T_n)$ の時間間隔が任意の確率分布 $F(x)$ に従うとすると、確率点過程 $\{N(t), t \geq 0\}$ をトレンド再生過程 (TRP($F(x), \lambda(t)$))) と呼ぶ。本稿では、確率分布関数 $F(x)$ を TRP の再生分布と呼び、 $\lambda(t)$ をトレンド関数と呼ぶ。定義より、トレンド再生過程は非同次ポアソン過程 (NHPP) 及び再生過程 (RP) を特殊な場合として含んでおり、それぞれ $\text{TRP}(1 - e^{-x}, \lambda(t))$ 及び $\text{TRP}(F(x), 1)$ で表現することが出来る。また、 $\text{TRP}(F(x), \lambda(t))$ の表記方法は一意ではなく、任意の定数 $c > 0$ に対して、 $\text{TRP}(F(x), \lambda(t))$ 及び $\text{TRP}(F(cx), \lambda(t)/c)$ が同じトレンド再生過程を表す性質を持つ [8]。

任意の確率点過程に対する条件付き強度関数は

$$\xi(t) = \lim_{\Delta t \rightarrow 0} \frac{P(\text{failure in } [t, t + \Delta t] \mid H_{t-})}{\Delta t} \quad (1)$$

と定義される。一般的に、条件付き強度関数 $\xi(t)$ は時刻 t より以前の確率点過程 $\{N(t), t \geq 0\}$ の事象履歴 H_{t-} に依存し、RP の場合、条件付き強度関数は $\xi(t) = z(t - T_{N(t-)})$ によって与えられる。ここで $z(x)$ は再生分布 $F(x)$ の故障率関数を表し、再生密度関数 $dF(x)/dx = f(x)$ を用いて $z(x) = f(x)/(1 - F(x))$ と定義される。一方で、NHPP の場合には条件付き強度関数は確率的ではなく、 $\xi(t) = \lambda(t)$ で表される。これに対して、 $\text{TRP}(F(x), \lambda(t))$ の条件付き強度関数は

$$\xi(t) = z(\Lambda(t) - \Lambda(T_{N(t-)}))\lambda(t) \quad (2)$$

で与えられる [8]。

今、 n 個の故障時刻データ $t_1, t_2, \dots, t_n$ が観測されたとすると、条件付き強度関数 $\xi(t)$ を持つトレンド再生過程の対数尤度関数は次のように与えられる。

$$LLF = \sum_{i=1}^n \left\{ \log(z(\Lambda(t_i) - \Lambda(t_{i-1}))) + \log(\lambda(t_i)) - \int_0^{\Lambda(t_i) - \Lambda(t_{i-1})} z(v) dv \right\}. \quad (3)$$

同様に、再生密度関数 $f(x)$ を用いて表現すれば、

$$LLF = \sum_{i=1}^n \left\{ \log(f(\Lambda(t_i) - \Lambda(t_{i-1}))) + \log(\lambda(t_i)) \right\} \quad (4)$$

となる。

3 再生密度関数のノンパラメトリック推定量

3.1 カーネル密度推定量

本稿では、トレンド関数が既知で、かつ再生分布が未知である場合において、カーネル密度推定量に基づいた TRP のノンパラメトリック推定法を提案する。確率分布関数 $F(x)$ からの i.i.d. 標本 $x_1, x_2, \dots, x_n$ が与えられている場合、カーネル密度推定量は以下のように定義される。

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_{h,i}(x). \quad (5)$$

ここで、 $K_{h,i}(x) = (1/h)K\{(x - x_i)/h\}$ であり、 $h(>0)$ はバンド幅と呼ばれる任意パラメータである。また関数 $K(x)$ はカーネル関数と呼ばれ、一般的に $\int K(u)du = 1, \int uK(u)du = 0, \int u^2K(u)du < \infty$ を満たす関数として定義される。本稿では、以下の式で表されるガウスカーネル関数を用いる。

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \quad (6)$$

カーネル関数に基づく密度関数の推定においては、カーネル関数の種類よりもむしろバンド幅の推定が重要であることが既に知られている。

ここでは、バンド幅を決定する手法として最小二乗クロスバリデーション法 (LSCV 法) 及び対数尤度クロスバリデーション法 (LLCV 法) を用いる。クロスバリデーション法に基づくバンド幅推定手法では、故障時刻列データを学習用データと検証用データに分割する。つまり、 $x_1, x_2, \dots, x_n$ から i ($= 1, 2, \dots, n$) 番目のデータの一つずつ抜き取ることで、 $n-1$ 個のデータからなる n 組の学習用データを生成する。任意の時刻 t における真の再生密度関数 $f(x)$ に対するカーネル密度推定量 $\hat{f}(x)$ の積分二乗誤差は、

$$\text{ISE}(h) = \int_0^{x_n} (\hat{f}(x) - f(x))^2 dx \quad (7)$$

と定義される。LSCV 法は、 h と独立な項を除いた場合、 $\text{ISE}(h)$ を最小化するバンド幅が以下の式を最小化するバンド幅と等価になる事実に基づいて推定を行う。

$$\text{LSCV}(h) = \int_0^{x_n} \hat{f}_h(x)^2 dx - 2 \sum_{j=1}^n \hat{f}_{h,j}(x_j). \quad (8)$$

一方、LLCV 法ではクロスバリデーション法を応用して尤度関数を最大化するようなバンド幅を求める。具体的には以下の式を最大化するバンド幅を推定する。

$$\text{LLCV}(h) = \sum_{i=1}^n \log \hat{f}_{h,j}(x_j). \quad (9)$$

ここで、式 (8) 及び式 (9) に対して、

$$\hat{f}_{h,j}(x) = \frac{1}{n-1} \sum_{i=1, i \neq j}^n K_{h,i}(x) \quad (10)$$

である。

3.2 トrend再生過程への応用

TRP の場合、 x_i は累積トレンド関数 $\Lambda(t)$ に依存する。つまり、累積トレンド関数の形状が特定されない限り、 $x_i = \Lambda(t_i) - \Lambda(t_{i-1})$ ($i = 1, 2, \dots, n$) を計算することが出来ない。ただし、 $\Lambda(0) = 0$ である。TRP

に対するカーネル密度推定量を計算する際、累積トレンド関数の変化に伴って時間スケール変換後の時間間隔 x_i もまた変化するという性質を考慮する必要がある。本稿では

$$\Lambda(t) = \alpha t^\beta, \quad (\alpha, \beta > 0) \quad (11)$$

で表されるベキ乗型の累積トレンド関数を仮定する。ここで、 α 及び β はトレンド関数 $\lambda(t) = \alpha \beta t^{\beta-1}$ のパラメータを表す。TRP の非一意性により、 $\text{TRP}(F(x), \alpha \beta t^{\beta-1})$ は $\text{TRP}(F(\alpha x), \beta t^{\beta-1})$ と同じ最大対数尤度を持つことから、計算を簡単化するために冗長なパラメータを $\alpha = 1$ と設定することが出来る。

本稿では、トレンド関数のパラメータ β 及びカーネル推定量のバンド幅 h を同時に推定することを考える。しかしながら、Lindqvist [9] の推定アルゴリズムと同様に、 β の初期パラメータに対してバンド幅 h を推定し、その後 β のみを更新していくようなアルゴリズムでは収束が保障されない。同様に、カーネル推定量を用いる場合、バンド幅 h を小さくすれば尤度が大きくなる性質があるため、単純に尤度関数を最大化する h を最尤推定量として求めることは不可能である。そこで本稿では、次のような繰り返し手法を提案する。まず計算の簡単化とアルゴリズムのパラメータ収束を保障するために、故障時刻データ t_i ($i = 1, 2, \dots, n$) を正規化した時刻列データ $\tilde{t}_i = t_i/t_n \in (0, 1]$ を用いる。続いて、パラメータ β の初期値 $\beta^{(0)}$ を設定し、カーネル密度推定量を求めるために、 $x_i(\beta^{(0)}) = \Lambda(\tilde{t}_i; \beta^{(0)}) - \Lambda(\tilde{t}_{i-1}; \beta^{(0)})$ ($i = 1, 2, \dots, n$) を用いてバンド幅を推定する。ここで、 $\Lambda(\tilde{t}_i; \beta^{(0)})$ は初期パラメータ $\beta^{(0)}$ を持つ累積トレンド関数である。バンド幅推定手法としては 3.1 節で紹介した LLSV 法もしくは LLCV 法を用いる。続いて、得られたバンド幅 $h^{(0)}$ 及びカーネル密度推定量 $\hat{f}_{h^{(0)}}(x)$ を使って、以下の問題を解くことで TRP の尤度関数を最大化するトレンド関数のパラメータ β ($= \beta^{(1)}$) を推定する。

$$\max_{\beta} (LLF(\beta)) = \max_{\beta} \left(\sum_{i=1}^n \left\{ \log(\hat{f}_{h^{(0)}}(x_i(\beta))) + \log(\lambda(\tilde{t}_i; \beta)) \right\} \right). \quad (12)$$

ここで、 $\lambda(t; \beta)$ はパラメータ β を持つトレンド関数である。同様の繰り返しを連続する β の推定値の差が任意の閾値 ϵ より小さくなるまで続ける。

4 数値例

4.1 シミュレーション実験

提案したカーネル密度推定量に基づくノンパラメトリック推定手法の推定精度を、モンテカルロシミュレーションで評価する。ベキ乗型のトレンド関数 $\Lambda(t) = 100 \times t^2$ 及びワイブル再生分布 $F(x) = 1 - e^{-x^3}$ を持つ TRP を仮定し、一般修理下での擬似故障時刻データを生成する。図 1 で示す典型的な 4 つのサンプルパスを用いて、計算を行った。ここで、DS1 は真のモデルの平均値関数に最も適合したデータであり、DS2 及び DS3 はそれぞれ平均値関数より大きい若しくは小さい故障件数を示すデータである。また、DS4 として平均値関数付近で S 字曲線を示すデータを用いた。これら 4 つの特徴的なデータに対して、パラメトリック推定法（最尤法）及びノンパラメトリックな推定手法を適用した。

本稿では比較対象として、5 種類のパラメトリックな TRP (ベキ乗型強度関数を持つ NHPP (P-NHPP), ガンマ再生分布を持つ RP (G-RP), ワイブル再生分布を持つ RP (W-RP), ベキ乗型トレンド関数とガンマ再生分布を持つ TRP (PG-TRP) [2], ベキ乗型トレンド関数とワイブル再生分布を持つ TRP (PW-TRP, 真のモデル) [8]) と 6 種類のノンパラメトリックな TRP (減少型トレンド関数とガンマ再生分布を持つ TRP (DG-TRP) [10], 増加型トレンド関数とガンマ再生分布を持つ TRP (IG-TRP) [10], 減少型トレンド関数とワイブル再生分布を持つ TRP (DW-TRP) [8], 増加型トレンド関数とワイブル再生分布を持つ TRP (IW-TRP) [8], ベキ乗型トレンド関数と減少型再生分布を持つ TRP (PD-TRP), ベキ乗型トレンド関数と増加型再生分布を持つ TRP (PI-TRP)) の 11 種類の統計モデルを仮定する。また、本稿で提案したベキ乗型トレンド関数

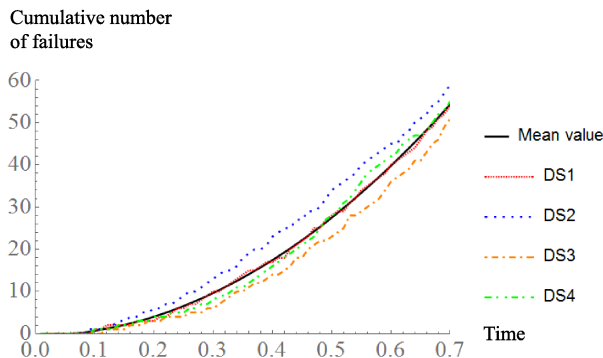


図 1: シミュレーション用データセット.

とカーネル再生分布を持つトレンド再生過程を PK-TRP と呼ぶ. TRP は NHPP 及び RP を特別な場合として包括する. 具体的には, PG-TRP 若しくは PW-TRP は, ガンマ分布若しくはワイブル分布の形状パラメータを 1 とした場合, P-NHPP に帰着される. また, PG-TRP 及び PW-TRP は, ベキ乗型トレンド関数の全てのパラメータを 1 とした場合, それぞれ G-RP 及び W-RP に帰着される.

図 2~図 9 は, 各データセットに対して提案手法を適用した場合の条件付き強度関数の推定結果を示したものであり, それぞれバンド幅推定手法として LSCV 法若しくは LLCV 法を用いた結果を表す. ここでは, 真のモデルである PW-TRP を用いて推定した結果を比較の対象として図示した. これらの結果から, LSCV 法を用いた場合の推定結果は真のモデルである PW-TRP と大きく異なる推定結果を示しているが, LLCV 法を用いた場合は PW-TRP と非常に近い推定結果を示していることが分かる.

更にデータへの適合性を定量的に測るため, 適合性の指標として最大対数尤度 (MLL) を用いた. 表 1 は比較対象を含めた 13 種類の統計モデルのデータ適合性を表したものである. 4つのデータセット全てにおいて, LLCV 法に基づく PK-TRP が真のモデルである PW-TRP に最も近い MLL の値を示していることが分かる. このことから, DS1~DS4 のように異なる特徴を持つデータに対しても, LLCV 法に基づく PK-TRP は真の統計モデルに近い推定結果を示すことが出来, 再生分布が完全に未知な場合において適用可能な上に優れた表現能力を持っていると言える.

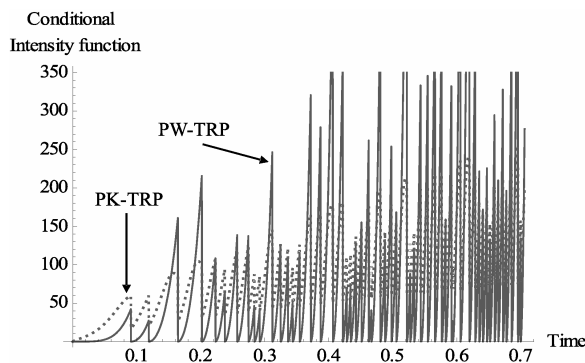


図 2: 条件付き強度関数の推定 (LSCV 法/DS1)

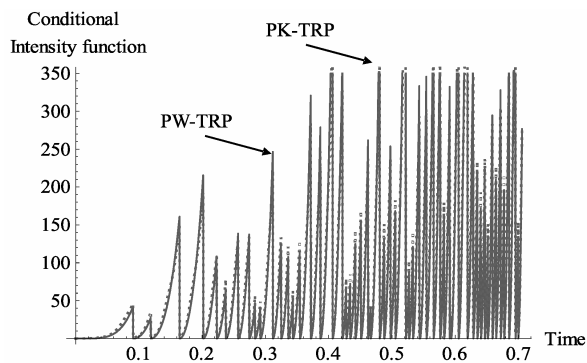


図 3: 条件付き強度関数の推定 (LLCV 法/DS1)

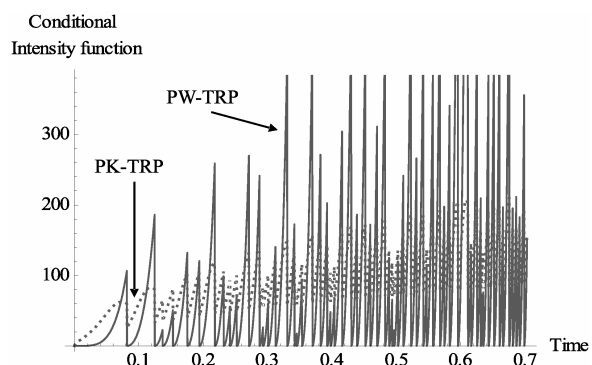


図 4: 条件付き強度関数の推定 (LSCV 法/DS2)

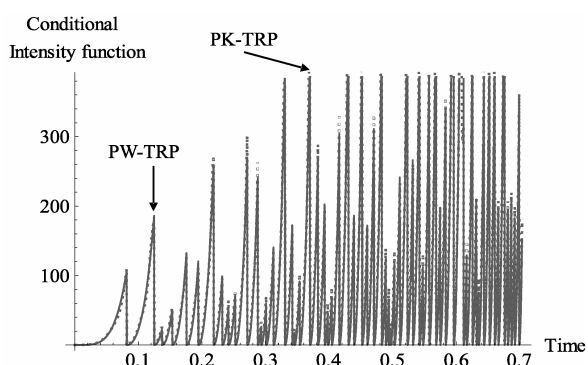


図 5: 条件付き強度関数の推定 (LLCV 法/DS2)

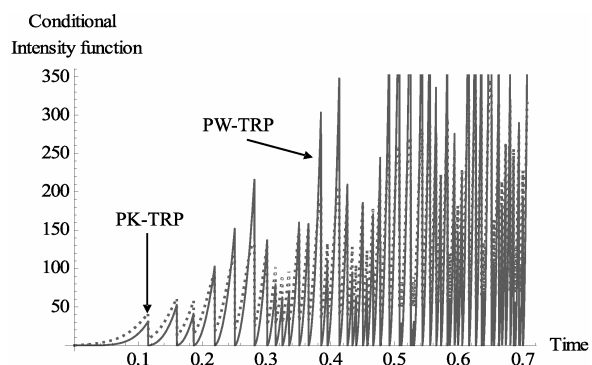


図 6: 条件付き強度関数の推定 (LSCV 法/DS3)

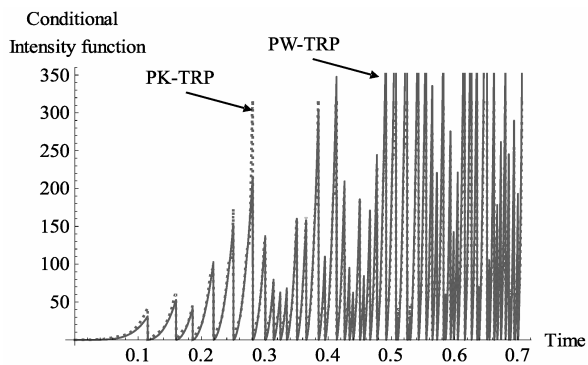


図 7: 条件付き強度関数の推定 (LLCV 法/DS3)

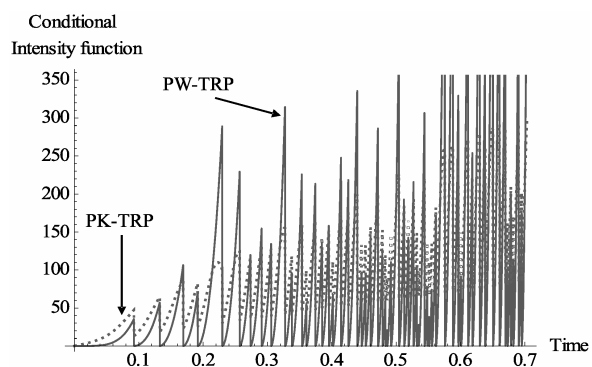


図 8: 条件付き強度関数の推定 (LSCV 法/DS4)

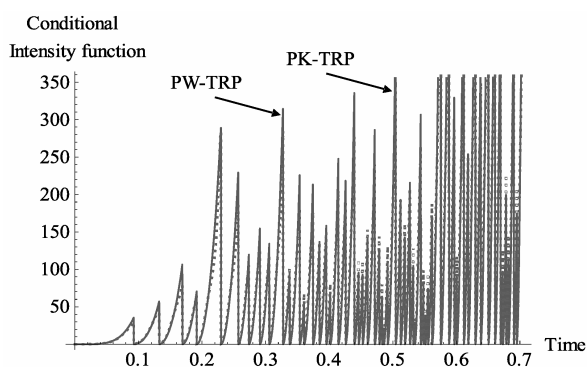


図 9: 条件付き強度関数の推定 (LLCV 法/DS4)

表 1: 適合性評価

Model	DS1	DS2	DS3	DS4
P-NHPP	196.33	216.27	191.66	202.49
G-RP	191.10	218.56	178.30	194.45
W-RP	187.52	212.50	176.47	191.26
PG-TRP	225.83	252.71	215.41	232.02
PW-TRP	227.43	253.23	217.52	232.51
DG-TRP	191.10	218.56	178.30	194.45
IG-TRP	225.86	254.44	215.40	233.72
DW-TRP	187.52	212.50	176.47	191.26
IW-TRP	232.11	259.19	221.92	237.54
PD-TRP	196.33	216.27	191.66	202.49
PI-TRP	253.57	270.69	245.86	259.09
PK-TRP(LSCV)	209.83	228.45	207.45	216.86
PK-TRP(LLCV)	226.94	253.15	219.44	232.36

4.2 実データ解析

ディーゼルエンジンの故障（修理）時刻データに対して、PK-TRP を適用した例を示す。図 10 と図 11 は、それぞれ全データ数の 50% 及び 100% を用いた場合の実データと累積故障数の推定結果を図示したものである。実故障データに対して本稿で提案した PK-TRP を適用し、比較の対象としてシミュレーション実験で用いた PW-TRP の推定結果を図示した。これらの図より、LLCV 法に基づく PK-TRP を用いた場合の推定結果は、PW-TRP を適用した場合と非常に近い値を示していることが分かる。100% の場合に着目すれば、特にデータの初期と終盤において、実データに非常に近い推定結果となっている。また、LSCV 法に基づく PK-TRP では、実データより少な目に故障数を見積る楽観的な推定結果を表しており、いずれの場合においても特にデータの終盤においてその傾向が強いことが伺える。

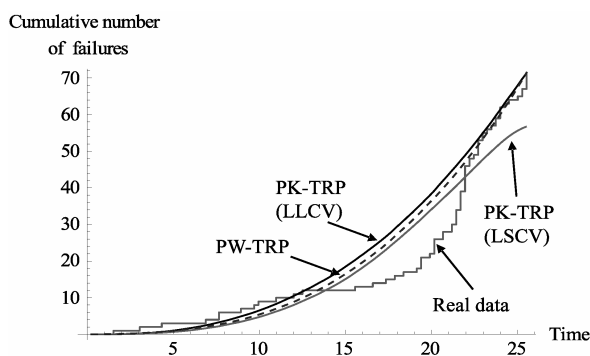


図 10: 累積故障数の推定 (50%点時刻)

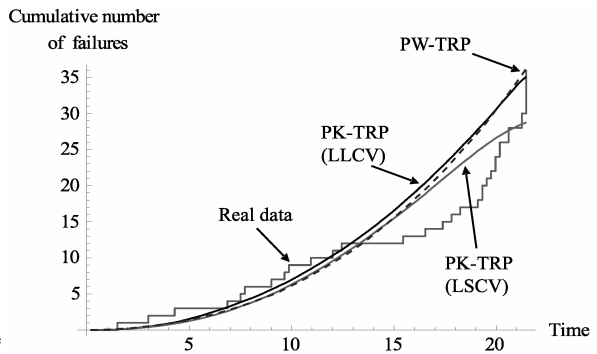


図 11: 累積故障数の推定 (100%点時刻)

5 まとめと今後の展望

本稿では、一般修理の下での修理系システムの故障の振る舞いをモデル化出来るトレンド再生過程に対して、再生分布をカーネル推定量によって表現したノンパラメトリック推定手法を提案した。具体的には、カーネル密度推定におけるバンド幅の推定手法として、最小二乗誤差に基づくクロスバリデーション法である最小二乗クロスバリデーション法 (LSCV 法) 及び尤度関数に基づくクロスバリデーション法である対数尤度クロスバリデーション法 (LLCV 法) の二つを用いた。更に、これらバンド幅推定法とトレンド関数に含まれるモデルパラメータを推定する最尤法を組み合わせた推定アルゴリズムを提案した。数値例を通じて、本稿で提案した LLCV 法に基づく PK-TRP は再生過程の関数形が未知な場合であっても、真のモデルと非常に近い推定結果を示すことが分かった。

今後の課題として、トレンド関数及び再生分布が共に未知である場合を想定したノンパラメトリック推定手法の提案が挙げられる。具体的には、トレンド関数や再生分布に対応する故障率関数の制約付きノンパラメトリック最尤推定量 (CNPMLE) とカーネル推定量を組み合わせた推定手法の考察を行なう予定である。

参考文献

- [1] H. Ascher and H. Feingold, *Repairable Systems-Modeling, Inference, Misconceptions and Their Causes*, Marcel Dekker, New York (1984).
- [2] R. Berman, "Inhomogeneous and modulated gamma processes," *Biometrika*, **68** (1), 143–152 (1981).
- [3] M. Brown and F. Proschan, "Imperfect repair," *Journal of Applied Probability*, **20** (4), 851–859 (1983).
- [4] D. R. Cox, "The statistical analysis of dependencies in point processes," *Stochastic Point Processes: Statistical Analysis, Theory and Applications* (Ed. P. A. W. Lewis), 55–66, Wiley, New York (1972).
- [5] C. Dorado, M. Hollander and J. Sethuraman, "Nonparametric estimation for a general repair model," *The Annals of Statistics*, **25** (3), 1140–1160 (1997).
- [6] K. Heggland and B. H. Lindqvist, "A nonparametric monotone maximum likelihood estimator of time trend for repairable systems data," *Reliability Engineering and System Safety*, **92** (5), 575–584 (2007).
- [7] M. Kijima, "Some results for repairable systems with general repair," *Journal of Applied Probability*, **26** (1), 89–102 (1989).
- [8] B. H. Lindqvist, G. Elvebakk and K. Heggland, "The trend-renewal process for statistical analysis of repairable systems," *Technometrics*, **45** (1), 31–44 (2003).
- [9] B. H. Lindqvist, "Nonparametric estimation of time trend for repairable systems data," *Mathematical and Statistical Models and Methods in Reliability* (Eds. V. V. Rykov, N. Balakrishnan and M. S. Nikulin), 277–288, Springer, New York (2010).
- [10] Y. Saito and T. Dohi, "Nonparametric estimation in software reliability assessment based on non-homogeneous gamma processes," *IEICE Technical Report*, **115** (47), 13–18 (2015).
- [11] Y. Saito and T. Dohi, "Nonparametric estimation method for trend renewal process based on failure rate function," *IEICE Technical Report*, **115** (379), 7–12 (2015).