

ベイズ推計による下水道機器の故障解析

中根 進

1. はじめに

下水道機器の故障を解析することにより、設備の年数の経過にもなう劣化や、劣化によって引き起こされる故障を予測することが可能となる。また故障解析の結果は、機器毎の点検周期、保全区分を設定するなど、設備保守管理の基本になるものである。

下水道機器の故障は、故障内容、停止時間などによって定義する必要があるが、一般に故障の頻度が少なく、個々の機器について故障解析することは難しい。文献¹⁾では寿命分布の確率密度関数や故障率を推定するには、少なくとも50個以上のデータを必要とするとしている。この前提になっているのは、観測データから信頼度値や確率密度分布値を求め、最小二乗法などを使って、それらの値を何らかの関数(ワイブル分布など)で近似する手法を使うためである。

本小文では、少ない故障情報(故障回数)の下水道機器に対して確率モデルを作成し、ベイズ推計により機器の故障回数を推計(シミュレーション)し、その推定故障回数を故障率に変換することによって、故障率の点推定と区間推定を行う手法を提案する。

2. 下水道機器の故障率

下水道機器の多くは、故障部を修理して、使用を続ける機器で、解析上「修理系」の機器である。筆者の調査した下水道機器は、故障実績が少ないことから、他の処理場の同種機器を含めてハザード解析を行った。この筆者の調査範囲の解析結果²⁾では、ワイブル分布の形状パラメータは $m \approx 1$ 、または $m < 1$ であり、故障率は経過年数によらずほぼ一定、または減少する形となった。また文献^{3)、4)}では、ドレニック(Drenick)の定理として、指数分布に基づき機器の故障解析がなされている。図-1に終沈汚泥掻寄機(チェーンフライト式)の故障解析事例²⁾を示す。

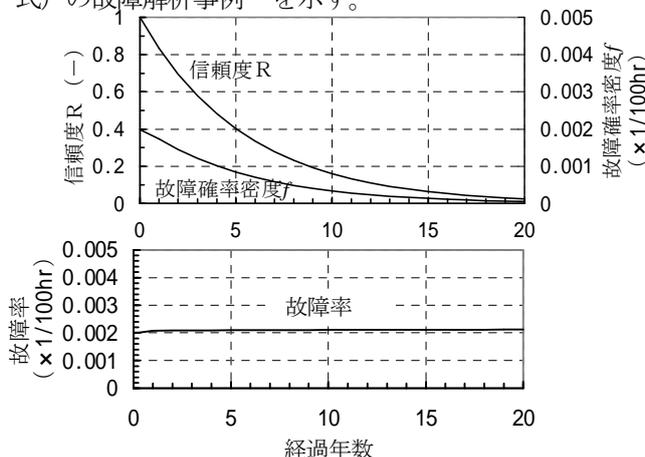


図-1 故障解析事例(終沈汚泥掻寄機)

下水道設備の維持管理記録は、日報ベースで記録され、月報や年報としても残されているが、月報や年報になるに従い情報が集約されてしまう。年数が経過すると、これら維持管理記録も年報の形でしか残っていないことが多く、機器の故障内容や故障時期・時刻も不明になっていることもあり、故障解析することを難しくしている。

そこで、少ない故障情報のうち故障回数のみ着目し、故障確率密度が指数分布する機器を前提に故障率の確率モデルをつくり、故障解析を行って故障率の点推定と区間推定を行う。

3. 確率モデル

3.1 確率モデルの考え方

事象を確率分布 y で表すことはよく知られている。その事象を観測や調査することで、事象の一部である標本を得ている。

観測や調査では、調査観測の器材能力、期間、資金などの制約を受けるため、得られる標本は、事象に対し多くの情報が失われている。また、調査観測者の持つ能力、知識などによってもばらつきを持つ標本となる。

事象を確率分布 y で表すことにより、情報が少なからずある調査・観測データ(標本)も、事象の確率分布 y の一部として表すことができる。

調査観測データからその事象の確率分布 y の形を推定する手法の1つに確率モデルの考え方がある。

確率分布 y は、パラメータ(例:正規分布であれば、平均 μ と標準偏差)で記述することができるので、そのパラメータ値を調査・観測データから見つけることをパラメータ推定という。そして多くの場合、調査・観測データに基づいて最小二乗法などで事象のパラメータを推定している。この手法は、観測データを多く必要とし、観測データが少ないとばらつきを表現することが難しい。

事象の確率分布の形を確率分布関数(確率密度関数)の組み合わせで表現したものを確率モデルといい、調査・観測データの形をうまく説明できるような確率分布関数とそのパラメータを使って確率モデルを表す。

4. ベイズの定理とベイズ推計

4.1 ベイズの定理

確率モデルのパラメータの不確実性を確率分布として表す。

調査観測データ y が得られれば、このデータ y で条件付けすることにより、未知パラメータに関する条件付き確率分布 $p(\lambda | y)$ は、ベイズの定理を使って (1) 式で表される。

$$p(\lambda | y) \propto p(y | \lambda) p(\lambda) \dots \dots (1)$$

一般に上式、左辺と右辺の各項を次のようにいう。

事後分布 $p(\lambda | y)$ = 尤度 $L(y | \lambda)$
 \times 事前分布 $p(\lambda)$

尤度関数 $L(y | \lambda)$ には調査観測データ y を入れ、事前分布 $p(\lambda)$ に調査・観測者の知見や先例のパラメータに対する分布とする。

4.2 階層ベイズ

4.1 の (1) 式はパラメータが 1 つの場合であるが、今回利用する後述の解析モデルでは、故障率をパラメータに利用するため、(1) 式を拡張する。

$$p(\lambda, \theta | y) \propto L(y | \lambda, \theta) p(\lambda | \theta) p(\theta) \dots \dots (2)$$

上式の左辺 $p(\lambda, \theta | y)$ は、条件付確率により下式となる。

$$p(\lambda, \theta | y) = p(\lambda | \theta, y) \cdot p(\theta | y)$$

上式右辺第 2 項は、下式となる。

$$p(\theta | y) = \frac{p(\lambda, \theta | y)}{p(\lambda | \theta, y)}$$

上式、右辺の分子に (2) 式を代入すると (3) を得る。

$$p(\theta | y) \propto L(y | \lambda) p(\lambda | \theta) p(\theta) \dots \dots (3)$$

ここで、調査観測データ y は、故障回数 x と調査期間 T とし、パラメータ λ には、 $\lambda = (x, T)$ を導入 (超パラメータ) する。

4.3 ベイズ推計

ベイズ推計は、故障回数 x のデータを使って尤度関数 $L(x | \lambda)$ を求め、何らかの事前分布 $p(\lambda)$ を使って (2) 式の事後分布 $p(\lambda | x)$ からサンプル (λ, θ) を得る。そのサンプル (λ, θ) を尤度関数 $L(x | \lambda, \theta)$ に代入して x のサンプルを得る。これを繰り返して n の標本を得て頻度 (確率) 分布をつくり、 n の点推定と区間推定を行う。具体的な分布の更新手順については 5.3 に示す。

4.4 使用ソフト

ベイズ推計する際のサンプリング手法の 1 つにマルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo methods: MCMC 法) がある。この MCMC 法の代表的なアルゴリズムであるギブス・サンプ

ラーを使ってシミュレーションする計算ソフトに WinBUGS という Windows 用のフリーソフト (09.7 現在 Ver1.4) がある。

WinBUGS は、ベイズ推計の尤度関数と事前分布の関数を記述すれば、条件付事後分布の計算は、WinBUGS 側で実行するソフトであり、今回の解析にはこのソフトを使った。

5. 解析モデル

5.1 終沈汚泥掻寄機解析データ

解析モデルに使用する収集データの構成を図-2 に示す。

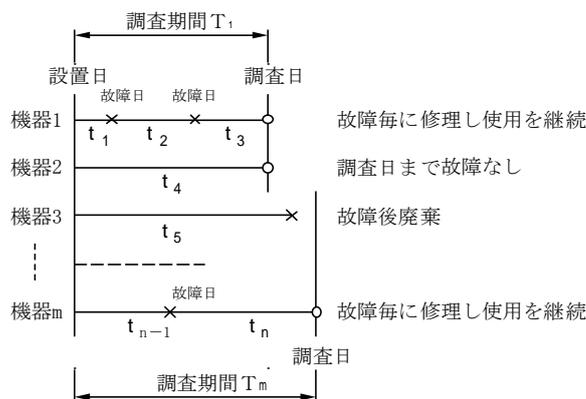


図-2 解析データの構成

図中の t_1, t_2, \dots は、故障日の間隔 (時間) であるが、この間隔は利用せず、今回の解析では故障回数のみを使用する。機器の設置日から調査日までの各機器の運転時間を調査期間 T_i とする。

調査期間 $T_i =$ 延べ運転時間 ($\times 100$ hr)

終沈汚泥掻寄機は、前掲図-1 のように故障率が一定を示しているため、故障確率密度のワイブル分布を指数分布と置き換える。

解析する終沈汚泥掻寄機データを表-1 に示す。

表-1 終沈汚泥掻寄機解析データ

処理場別 機器名	設置年	第 1 回調査 (1990.4.1)			第 2 回調査 (2000~04.4.1)		
		運転時間 $\times 100$ hr	故障 回数	故障率 $\times 1/100$ hr	運転時間 $\times 100$ hr	故障 回数	故障率 $\times 1/100$ hr
T1	1983/4/1	614	1	0.001629	1,527	5	0.0032744
T2	1983/4/1	614	2	0.003257	1,527	3	0.0019646
T3	1983/4/1	614	1	0.001629	1,526	3	0.0019659
U1	1986/7/1	592	1	0.001689	1,514	2	0.001321
U2	1986/7/1	592	1	0.001689	1,513	3	0.0019828
H I	1982/4/1	614	4	0.006515	1,815	6	0.0033058
H1	1986/4/1	614	0	0	1,625	2	0.0012308
H2	1986/4/1	614	0	0	1,625	1	0.0006154
計		4,868	10	0.002054	12,672	25	0.0019729

表中の第 2 回調査は、筆者らの実際に行った調査を示すが、第 1 回調査は、解析データ増加と解析精度の

関係を調べるために調査日を仮想し、その仮想した調査日までの実績故障日から故障回数を算出した。

機器設置日からある調査日までの調査期間内の故障率 λ_i は、指数分布することを前提に (4) 式で表す。

$$\text{故障率 } \lambda_i = \frac{\text{故障回数 } x_i}{\text{調査期間 } T_i} \dots\dots\dots (4)$$

(4) 式で算出した故障率を前掲表-1にあわせて示す。ここで、仮想した第1回調査では故障実績が少なく、故障回数が1回の終沈汚泥掻寄機が4基、故障が無い(0回)終沈汚泥掻寄機はH1, H2の2基あった。

5.2 終沈汚泥掻寄機解析モデル

故障解析のために確率モデルを次のように設定する。故障回数 x_i は、調査によって得るが、(4) 式を書き換える。調査期間 T_i は、ゼロから調査時までの機器の延べ運転時間である。

$$\text{故障回数 } x_i = \text{故障率 } \lambda_i \times \text{調査期間 } T_i \dots (5)$$

故障回数 x を文献⁵⁾に示されるポアソン分布で表す。

$$p(x) = \frac{\mu_i^x \exp(-\mu_i)}{x!} \dots\dots\dots (6)$$

ここに、 μ_i : ポアソン分布のパラメータで各機器の故障回数 x の平均を表す。

ポアソン分布の形状例を図-3に示す。

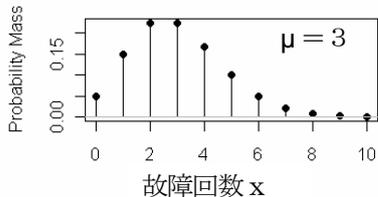


図-3 ポアソン分布の一例

(5) 式の故障回数は調査期間中の平均回数であり、(6) 式のパラメータ μ に代入し、(7) 式とする。

$$L(x_i, T_i | \lambda) = \frac{(\lambda_i \cdot T_i)^{x_i} \exp(-\lambda_i \cdot T_i)}{x_i!} \dots\dots (7)$$

ここに、 i : 機器

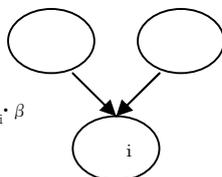
x_i : 各機器の調査期間内の故障回数 (回)

λ_i : 各機器の故障率 (回/100hr)

T_i : 調査期間 (×100hr)

故障率 λ_i は、ガンマ分布⁶⁾とする。ガンマ分布は、二つのパラメータ (平均 α と分散 β) を使って次式で表す。

$$p(\lambda_i | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \lambda_i^{\alpha-1} \cdot e^{-\lambda_i \cdot \beta}$$



ここに、 α : 故障率の平均

β : 故障率の分散

$\Gamma(\cdot)$: ガンマ関数

George ら (1993)⁷⁾ は、超パラメータ (故障率の平均)、 (故障率の分散) に対して次の指数分布とガンマ分布を事前分布に設定した。

$$p(\alpha) = \frac{1}{a} \cdot e^{-\frac{1}{a} \cdot \alpha}$$

ここに、 a : 平均

$$p(\beta) = \frac{c^b}{\Gamma(b)} \cdot \beta^{b-1} \cdot e^{-\beta \cdot c}$$

ここで、 b : 平均

c : 分散

以上の式と分布関数を取りまとめて確率モデルを図-4とし、解析モデルとする。

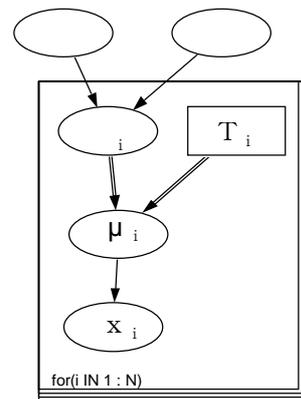


図-4 今回の解析モデル

ここに、

→ : ノードを確率分布で接続 (確率分布による依存)

⇒ : ノードを関数で接続 (関数による依存)

□ : 常数を表すノード

○ : ノード

i : 解析機器

N : 解析機器数

このモデルに対しギブス・サンプラーを使って乱数を発生させ、必要な事後分布を推定する。

以下に示す解析では、超パラメータ a と c の事前分布の初期値には、次式に示す平均、分散を使った。

$$p(\alpha) = \frac{1}{1} \cdot e^{-\frac{1}{1} \cdot \alpha}, \quad p(\beta) = \frac{1.0^{0.1}}{\Gamma(0.1)} \cdot \beta^{0.1-1} \cdot e^{-\beta \cdot 1.0}$$

5.3 MCMC 法のギブス・サンプラーの概説

WinBUGS 中の分布の更新手順を筆者の理解の範囲で概念を示す。

- ①パラメータ(,)の分布を仮定し、その分布からサンプリングした多数の(,)値と機器 i の故障率 λ_i を使って故障率 λ_i の分布を作る。
 - ②故障率 λ_i の分布から故障率 λ_i 値を多数サンプリングする。
 - ③このサンプリングした故障率 λ_i 値とデータの調査期間 t_i から故障回数 x の分布を作成する。
 - ④故障回数 x の分布の平均値とデータの故障回数 x_i を比較する。
 - ⑤この故障回数 x の分布と 1 つ前に試行した分布の比を採択確率として、パラメータ(,)の分布を少し変える。
 - ⑥①以降を繰り返し試行し、設定の試行回数で計算を止める。
- ここまで計算した後にパラメータ(,)と各機器 i の故障率 λ_i の分布を得る。

6. 解析結果

表-1 に示す各処理場の終沈汚泥掻寄機の故障回数を使った解析結果を以下に示す。

6.1 計算過程

表-1 に示す終沈汚泥掻寄機解析データに対し、WinBUGS を使って MCMC 法により、各母数に対して 11,000 回の繰り返し計算を行い、そのうち初期値の影響が残る恐れのある最初の 1,000 回を捨て(バーニン)、残りの 10,000 回から標本を取り出す。今回は試行 1 回で 1 個の標本をサンプルしたので 10,000 個の標本を発生させた。

各超パラメータと故障率の標本発生状況の試行回数 1,001 回から 11,000 回の一例を図-5 に示す。

試行回数やバーニン回数は、ギブス・サンプラー(マルコフ連鎖)の収束具合で設定する。収束の具合は図-5 の標本値の変動が一定になることを確認することであり、サンプル前の時点の標本値との自己相関係数をさまざまな間隔で計算した自己相関関数図を作って確認する方法やバーニン期間の標本を捨てた残りの標本の前半部分と後半部分に対し、それぞれの平均 E と分散 V を計算し Geweke の Z 値を使い、有意検定する方法などがある。

ここでは、WinBUGS で出図される図-6 の自己相関係数図で標本値が比較的近いサンプリング間隔において 0 近辺に収束しているかどうかを見ることによって、マルコフ連鎖の収束判定⁸⁾する。

直接、試行回数やバーニン回数を決定するものではないが、標本から作成した確率密度分布が不変分布(定常分布)に至っているかどうか確認するため、WinBUGS には、パラメータ毎に MC エラー(モンテカルロエラー)が出力される。また、WinBUGS によるシミュレーション結果は、信頼区間を入力すると、標本から得られる

事後統計量として確率分布図や点推定値(中央値)と区間推定値が表として出力される。

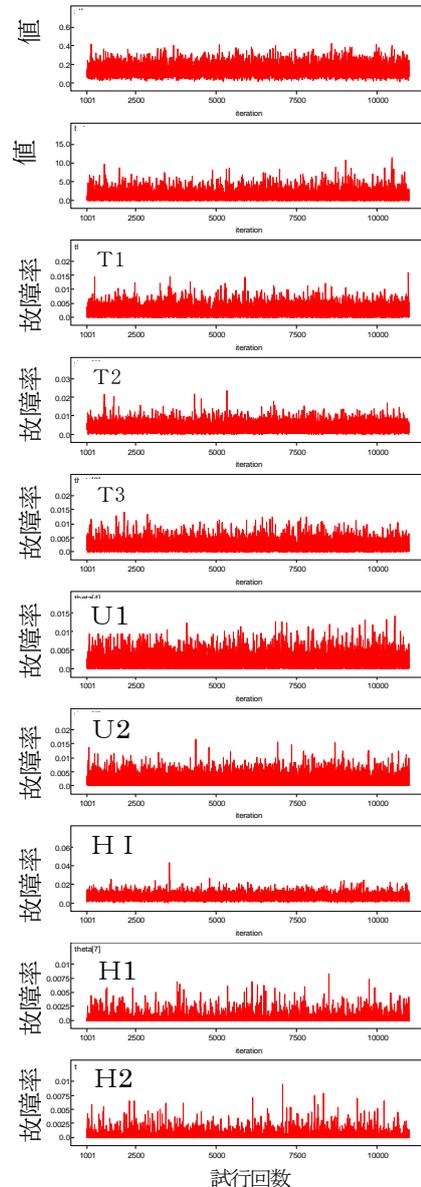


図-5 標本発生状況

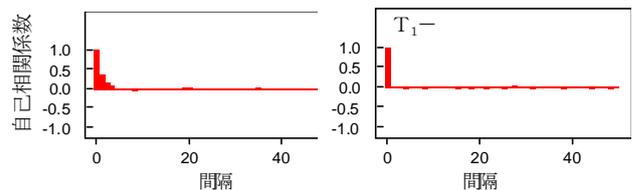


図-6 超パラメータ と故障率 の自己相関関数図の一例

図-5 の標本から作成した確率密度分布の一例として超パラメータ λ の事後分布を図-7 に、機器 T2 の故障率 λ_{T2} の事後分布を図-8 に示す。

ても変化しないと考えられた。

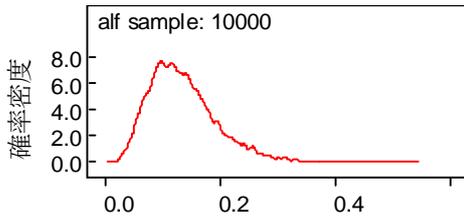


図-7 超パラメータ の事後分布

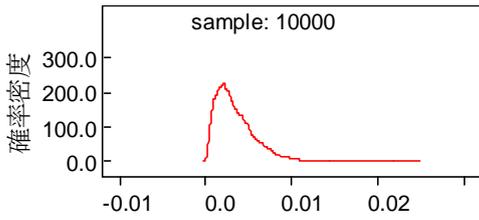


図-8 機器 T2 の故障率 の事後分布

6.2 計算結果 (事後統計量)

第1回調査と第2回調査時の標本 10,000 個から基本統計量を算出して、事後統計量として表-2, 3に示す。表-2, 3に前述(4)式で算出した故障率をあわせて示す。

表-2 第1回調査 事後統計量

処理場別 機器名	故障率 × 1/100hr	事後平均 i	標準偏差 sd	MCエラー	95%信頼区間		
					下限値2.5%	中央値	上限値97.5%
		0.1354 1.175	0.05783 1.165	7.75E-04 1.51E-02	0.04804 0.01431	0.1271 0.8353	0.2711 4.252
T1	0.001629	0.001858	0.001726	1.73E-05	6.712E-05	0.001343	0.006408
T2	0.003257	0.003436	0.002346	2.29E-05	0.0004815	0.00291	0.009262
T3	0.001629	0.001848	0.001739	1.85E-05	6.651E-05	0.001344	0.006422
U1	0.001689	0.001873	0.001741	1.73E-05	6.687E-05	0.001372	0.006514
U2	0.001689	0.001886	0.001774	1.84E-05	0.0000663	0.00138	0.006729
H I	0.006515	0.006689	0.003285	3.38E-05	0.001929	0.006184	0.01434
H1	0	0.000222	0.000608	5.90E-06	8.291E-21	4.75E-06	0.00211
H2	0	0.000216	0.00058	5.85E-06	3.041E-20	4.86E-06	0.001963
平均値	0.002051	0.002254			0.0003347	0.001818	0.0067185

表-3 第2回調査 事後統計量

処理場 名称	故障率 × 1/100hr	事後平均 i	標準偏差 sd	MCエラー	95%信頼区間		
					下限値2.5%	中央値	上限値97.5%
		0.1823 1.53	0.06744 1.339	6.57E-04 1.39E-02	0.07648 0.05311	0.1743 1.163	0.3357 4.973
T1	0.003274	0.003383	0.001486	1.59E-05	0.001125	0.003185	0.006872
T2	0.001965	0.002093	0.00116	1.16E-05	0.0004556	0.00188	0.004966
T3	0.001966	0.002081	0.001153	1.17E-05	0.0004654	0.001872	0.00482
U1	0.001321	0.001435	0.000968	9.49E-06	0.0002034	0.001218	0.003894
U2	0.001983	0.002095	0.00117	1.20E-05	0.0004537	0.001892	0.004934
H I	0.003306	0.003412	0.001372	1.54E-05	0.001279	0.003226	0.006566
H1	0.001231	0.001326	0.000895	9.99E-06	0.0001896	0.001125	0.003554
H2	0.000615	0.000726	0.000674	7.11E-06	2.777E-05	0.000524	0.002455
平均値	0.001958	0.002069			0.0005249	0.001865	0.0047576

表中のMCエラー(モンテカルロエラー)は、推定された各パラメータ(, ,)の事後平均の標準誤差(標準偏差 / √(サンプル数N))を示し、第1回調査では最大でも0.003%であり、第2回調査では0.001%となり、分布についてはこれ以上の試行回数し

6.3 考察

ベイズ推計で得られた表-2, 3の各機器の事後統計量を点推定値と区間推定値として図-9に示す。第1回調査時の区間推定値を上段に、第2回調査時を下段に示す。

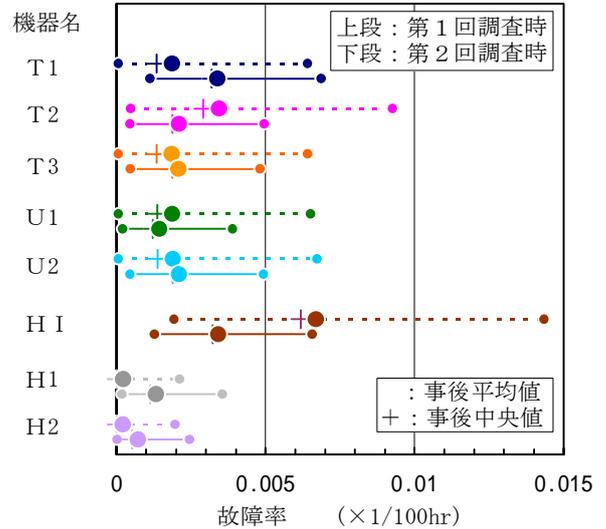


図-9 終沈汚泥掻寄機の調査時期と区間推定

第1回目の調査日は仮想した日にちであるものの故障率の区間推定幅は、第1回より故障データが増える第2回調査の方が狭くなっている。

機器H1, H2については、第1回調査で故障数がゼロであるため、故障データが得られた第2回調査の方が区間推定の幅が広がっている。

ベイズ推計では、第1回調査では故障数がゼロであっても、事前分布があれば、区間推定値が計算できる。リスク評価などでは区間推定の上限値が重要となることもあり、上限値が示されたことにより、故障数が観測されない機器についても、故障があった機器と同じように評価することができることを示している。

(4)式で算出した故障率は、事後平均値と95%信頼区間の事後中央値の間となった。

$$\text{事後中央値} \leq (4)\text{式の故障率} \leq \text{事後平均}$$

(4)式の故障率とベイズ推計で得られた故障率の事後平均の関係を図-10に示す。ベイズ推計の95%信頼区間の下限値2.5%、上限値も破線97.5%で示す。

図-10には参考に表-1の終沈汚泥掻寄機を同一機器としてハザード解析して得た故障率(図-1参照)も示す。

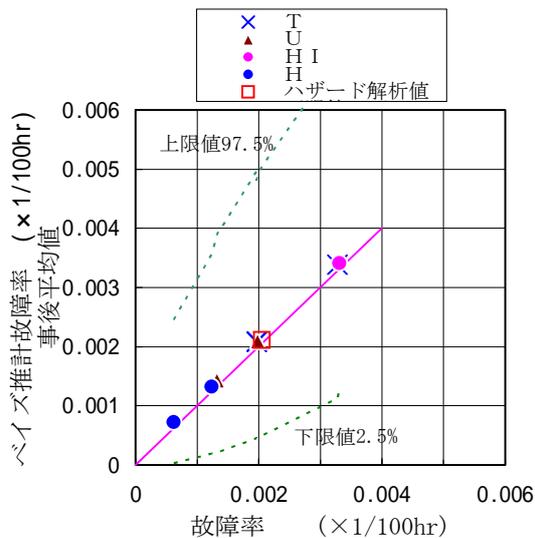


図-10 終沈汚泥掻寄機(4)式の故障率 とベイズ推計の故障率の関係

このベイズ推計結果から、故障確率密度が指数分布する機器では、ベイズ推計を行わなくても単純に(4)式で故障率を算出し、その故障率で日常の設備保守管理の指標とすることや、経過時間に対する信頼度(故障しない確率)に変換して設備の劣化状態を知ることができる。しかし、第1回調査時のように故障が無い場合には(4)式で故障率も算定できないし、ハザード解析もできないので、ベイズ推計は有用な故障解析手段となる。

第1回目で故障が0(ゼロ)であったH処理場のH1終沈汚泥掻寄機の信頼度を故障率の事後平均値と上限値から変換して図-11に示す。

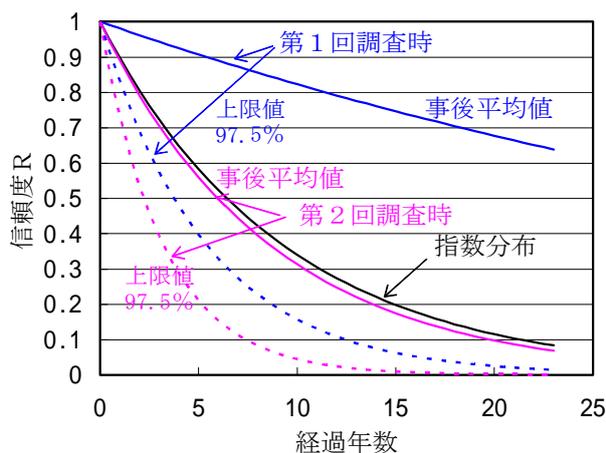


図-11 H1の経過年数と信頼度

実線は、事後平均値、破線は95%信頼区間の上限値から変換した信頼度曲線である。

図-11に示す故障確率密度が指数分布の信頼度曲線は、故障が発生した以降でないとは描けない曲線であ

り、日常の設備管理には利用できない時期がある。

その点、ベイズ推計を用いれば、故障率値の逆数を点検周期の目安としたり、故障率から信頼度R、また故障確率F(=1-R)を求め、その機器の経過年数までに故障する確率値を把握して、維持管理の優先順位に利用したり、95%信頼区間の上限値から故障確率Fを求め、機器のリスク管理に利用できる。

7. まとめと課題

下水処理場の機器は、故障回数が少ないため故障解析を難しくしているが、本確率モデルのベイズ推計により、故障回数が少なくても故障率の点推定や区間推定が可能になり、経過年数に対する信頼度曲線を得ることもできる。

故障回数がゼロの終沈汚泥掻寄機2基に対してもベイズ推計により区間推定することができた。

故障データが増えれば精度の向上も区間推定の幅が小さくなることで示すことができた。

ベイズ推計により我国あるいは自ら管理する下水道機器種別毎の事前分布の超パラメータ、分布を示すことができれば、日常の管理で観察した故障回数を使い、故障実績の少ない時点から将来得られるであろう故障率に近い値、分布が得られる。この故障率の中央値、区間推定値を使うことにより、機器の状態(経過年数による信頼度)、故障の予測(故障確率)を推定でき、設備管理の手助けができると考えられる。

今後、故障データを入手したい指数分布した終沈汚泥掻寄機以外の下水道機器を調査し、本小文で示した確率モデルの適用範囲や他の確率モデルを研究していきたい。

<参考文献>

- 1) 改訂版 信頼性工学入門 p.71 真壁 肇 編 日本規格協会
- 2) 下水処理場プラント機器の信頼性解析事例—機器の累積ハザード解析— 中根 進 下水道協会誌 論文集 2006.4月号
- 3) 1) p.84
- 4) 機械安全工学—基礎理論と国際規格— p.23 清水久二, 福田隆文 編著 2000 養賢堂
- 5) 1) p.123
- 6) 1) p.122
- 7) Pumps : conjugate gamma-Poisson hierarchical model WinBUGS14—Help—Example Volume I
- 8) マルコフ連鎖モンテカルロ法 p.40 豊田秀樹編著 朝倉書店