



通電式塗膜評価におけるデータのモデル探索

磯貝, 恭史
打田, 浩明

(Citation)

神戸大学大学院海事科学研究科紀要, 04 <商船・理工論編>:89-100

(Issue Date)

2007-07

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/80050009>

(URL)

<https://hdl.handle.net/20.500.14094/80050009>



通電式塗膜評価におけるデータのモデル探索

Model Detection of Enamel Rater Value Data

Kobe University

磯貝恭史, 打田浩明*

Takafumi ISOGAI, and Hiroaki UCHIDA*

(平成19年度4月6日 受付)

Abstract

The inside of a metallic can filled with some drink is usually coated with a macro-molecular film to suppress the progress of corrosion. The degree of its coating is evaluated by so-called Enamel Rater Value (shortly, ERV) of a quick electrical stress test. ERV is nonnegative-valued, and its small value is desirable for coating. Our objective here is to examine distributional shapes of ERV data sets coming from typical manufacturing processes of aluminium cans under control. The data sets are left-censored and roughly classified into two groups according to specifications of coating. But, they have a notable common feature that each distribution of them is anti J-shaped with a very high peak near the origin and also an extremely spread tail. Four statistical distributions, including the generalized Gamma, have been adopted to fit the data sets to examine their performance. The best one is a new power normal family introduced by one of the authors (Isogai [3]). The second one is the generalized Gamma, but it has a serious computational problem. Finally it is also shown that our new power normal family gives us clear information about distributional shapes of the ERV data..

(Received April 6, 2007)

1. はじめに

缶飲料などで多用されているアルミニウム（以下、アルミニウムをアルミと略記する）缶は、充填後の経時腐食を防ぐ目的で高分子塗膜が内面塗布される(西山[7])。塗膜としてエナメルが利用されていた時期から、この塗布状態の評価のために通電式の簡易検査法が採用され、エナメルレーター値 (Enamel Rater Value, 以下ERVと記す) が計測されてきた。ERVは1% 食塩水を電解液とし、缶体をアノードに、ステンレス棒をカソードとして、その間に6.2ボルトの電位差を4秒間与えた直後の瞬間的な電流値(mA)を指す。その値の大きさが内面塗膜の覆われていない塗膜欠陥（金属露出部）の個数と面積を検出すると考えられている。飲料充填後のアルミ缶は、もし塗膜欠陥が存在すれば、その部位から経時変化による孔食腐食が生じて、アルミ溶出を引き起こすことが分かっている。そのため、より小さなERV値が経時腐食によるアルミ溶出防止に有効であると考えられている。

ERV計測法の単純さに対して、その計測が塗膜欠陥のアノード溶解反応によるアルミ溶出を伴う複雑な現象を扱っているため、ERVそのものの物理的内容および電気化学的な内容についての定量的・定性的な評価を行った文献が全く見あたらない (ERVの物理化学的な批判検討については、西山他[8]参照)。また、製造現場ではERV計測が普通に行われているにもかかわらず、母集団分布についての定量的な評価がなされず、単に工程異常の発見という立場での評価が主であった。

管理状態下にある製造工程で造られたアルミ缶のERVデータの分布は、0値を伴う原点近くでの高いピークと非常に長い右裾が特徴的な反J字型をしている。その母集団分布の理論的な導出については、計測問題の立場からの研究 (Isogai & Uchida [1], Isogai [2]) があるだけで、ほとんど何も知られていない状況であったが、最近、著者達に

* ユニバーサル製缶株式会社 技術開発本部

よって、ERVの電気化学的な性質を調べるための実験計画を通して、理論的な分布の構築が試みられている(Isogai et al.[5])。しかし、そこで得られる理論分布は特殊関数の数値積分が現れてくるような非常に複雑なもので実用的には取り扱いにくく、簡便法が望まれる。そのため、この論文ではERVの分布形状の情報を与えるような取り扱い易い統計的モデルを採用して、それらのERVデータへの適合性に関して調べた比較検討結果を報告する。

取り上げた理論分布としては、Isogai & Uchida [1]とIsogai [2]で提案された二つの分布と一般化ガンマ分布 (Johnson & Kotz [6] 参照) , および変数変換を用いる新ベキ正規分布 (Isogai[3], 磯貝[4]) である。ERVデータを用いて、これら四つの分布族の性能を調査した結果、一般化ガンマ分布と新ベキ正規分布については、データへの適合性は十分にあり、また適合性には余り差がないことも判明した。ただ、一般化ガンマ分布は本質的に母数の推定に関して数値計算上の難点を抱えており、簡便法として新ベキ正規分布の適用が勧められる。

2. ERVデータに対する統計的モデル

1. データセット

アルミ缶は、缶胴と缶蓋の2ピースから構成されている。缶胴は、コイルシートを絞り・深絞り加工して、その側壁を元板厚の約1／3とし、洗浄、表面処理、印刷、外面・内面塗装後、ネッキング・フランジ加工することによって作られる。一方、缶蓋は、塗装したコイルシートから缶胴と嵌合できる形状にシェル成形し、その後飲み口部のスコアリング、タブをリベッティングすることにより加工される。このように、缶胴と缶蓋の内面塗膜への欠陥をもたらす要因が異なるため、ERVの母集団分布も違う可能性がある。従って、ここでは缶胴及び缶蓋について取り扱うこととした。

表1 缶胴と缶蓋のデータセットの標本サイズの構成

部位	データセット	閾値c	n (データ数)	n ₊ (正の値の個数)	n ₀ (0値の個数)
缶胴	A(高)	0.1	500	23	477
	A(低)	0.1	500	240	260
	B(高)	0.1	500	186	314
	B(低)	0.1	500	471	29
缶蓋	E(高)	0.01	1425	1425	0
	E(低)	0.01	480	454	26

高：腐食性の高い飲料用，低：腐食性の低い飲料用

表2 缶胴と缶蓋のデータセットの基本統計量

データセット	A(高)	A(低)	B(高)	B(低)	E(高)	E(低)
平均値	0.0628	8.6008	0.1188	4.1802	0.3553	1.0704
中央値	0	0	0	1.6	0.18	0.155
標準偏差	0.4323	13.537	0.2989	7.0199	0.7296	2.2678
歪度	9.8429	2.7699	6.0401	3.3782	11.749	6.4509
尖度	118.17	12.723	49.570	14.420	191.12	72.202
最小値	0	0	0	0	0.01	0
最大値	6.5	103.4	3.3	51.7	14.37	31.8

さて、ERVデータの母集団分布の形を探るため、缶胴と缶蓋について集めたデータの種類を表1に示す。缶胴については2つの工場(AとB)で生産されている腐食性の相対的に高い飲料用(「高」と略述)と低い飲料用のもの(「低」と略述)を対象とし、缶蓋については国内生産の腐食性の相対的に高い飲料用のもの(E(高))と米国製の腐食性の相対的に低い飲料用のもの(E(低))を採用した。缶胴及び缶蓋のERVデータは、生産中に系統サンプリング方式により抜き取り計測される。今回の調査対象は、工程でのセットアップ条件が変更されてない範囲を抽出しプールしたものである。

まず、表1を見て気づくことは、缶胴と缶蓋のERVの0値の現れ方の相違である。特に腐食性の相対的に高い飲料用の缶胴のERVの0値がきわめて多く、一方、缶蓋のERVの0値は非常に少ない。この中でA工場の腐食性の相対的に高い飲料用の缶胴のERVは大部分が0値であり、分布の探索には適さないと思われるが、工程としては代表的な管理状態であるので、一応取り上げておく。

また、ERVが0になっているデータを調べて判明したことであるが、ERVが小さな値の時には自動的に0と記録されてしまうように計測器が設定されていた。そのときの設定値を閾値cとして表1に載せている。

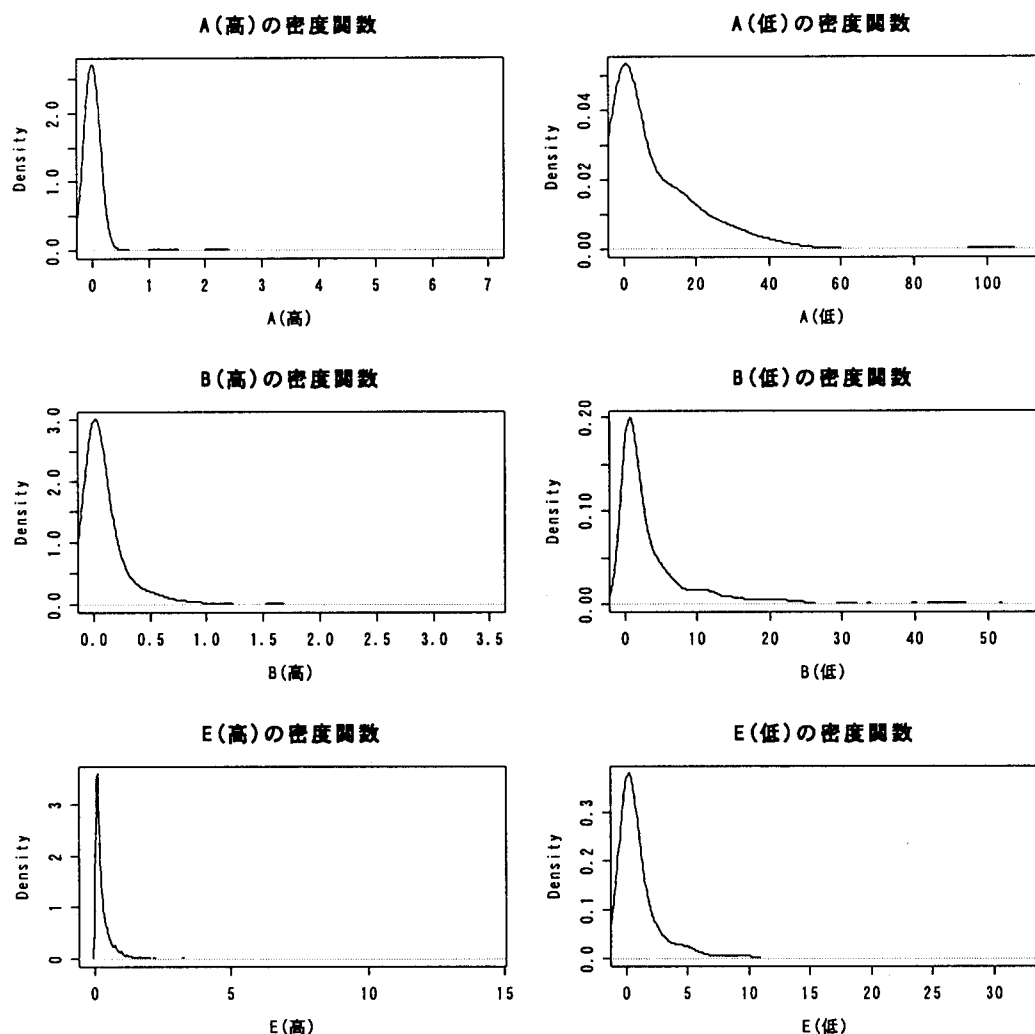


図1 缶胴・缶蓋のデータセットに関する推定された密度関数

図1および表2にデータセットの推定した密度関数と基本統計量を示す。すべてのデータセットが類似した分布形状を持っていることが判る。歪度と尖度の値が大きく、特に尖度の値が飛び抜けて大きいのが特徴的である。従って、分布の適合のためには0の近辺に適合するだけでなく、遠くの裾にも適合する分布を探さねばならない。

2. 分布の候補

1. 理論分布による接近法

さて著者達は, ERVが荷電粒子の電解液中を移動する状態を描写していると考えた. そして, その時の移動にかかる時間を $T(>0)$ とし, 内面塗膜の塗装の良さを示す変数を $W(>0)$ として, ERVが

$$(1)ERV = T \cdot W$$

$$(2)ERV = \frac{W}{T}$$

の二つのモデルのいずれかによって表されると考えた. 最初のモデル(1)は荷電粒子の速度を一定としたときに, 電流値ERVが荷電粒子の仕事量(速度 $\times T$)に比例するとし, 電流の流れを変数 W が制御していると考えた. モデル(2)は荷電粒子の動く距離が一定であるときには, 荷電粒子の運動量(質量 $\times$ 距離/ T)がERVに比例するとし, 電流の流れを変数 W が制御するものと考えた. さらに, 変数 T の分布として荷電粒子のブラウン運動で用いられる逆正規分布(Johnson & Kotz [6] 参照)を, 変数 W の分布として指数分布を仮定してERVの分布をそれぞれ導出した(詳細は, Isogai & Uchida [1], Isogai [2]参照). 分布の定義の形から, モデル(1)をProduct分布, モデル(2)をRatio分布と呼んだ.

以下では, 観測対象であるERVを確率変数 X で表し, その密度関数を $f_{\theta}(x)$ で, 分布関数を $F_{\theta}(x)$ で表すことにすれば, 二つの分布は

(1) Product分布

$$F_P(x) = 1 - \frac{1}{\sqrt{1+2x\varphi\beta}} \exp \left[-\varphi \left(\sqrt{1+\frac{2x}{\varphi\beta}} - 1 \right) \right] \quad (x \geq 0, \varphi > 0, \beta > 0)$$

(2) Ratio分布

$$F_R(x) = 1 - \exp \left[-\varphi \left(\sqrt{1+\frac{2x}{\varphi\beta}} - 1 \right) \right] \quad (x \geq 0, \varphi > 0, \beta > 0)$$

で与えられる.

Product分布とRatio分布は共に $x=0$ で最頻値を持つ反J字型の分布である. 両者は尺度母数 β と形の母数 φ を持つ. さらに両者とも, $\varphi \rightarrow \infty$ の時, 分布は平均 β の指数分布に近づき, $\varphi \rightarrow 0$ のときには, 分布は原点0に集中し, 指数分布よりも遙かに長い右裾を持つ分布に近づく.

一方, E(低)については米国の企業の研究者から, 私信で対数正規分布の可能性を示唆してもらっている. また, ERVデータの今までの調査結果から指数分布やワイブル分布も母集団分布の可能性があることが分かっている. そのため, ERVデータの母集団分布は, Product分布やRatio分布であるかということよりも遙かに広い分布族を含んでいると考えられる. 従ってワイブル分布や対数正規分布などの異なる分布族を含む族として, 一般化ガンマ分布族(Johnson & Kotz [6] 参照)がERVデータの統計的モデルの一つの候補となる.

一般化ガンマ分布族の分布関数 $F_{\theta}(x)$ は

(3) 一般化ガンマ分布

$$F_G(x) = \frac{1}{\Gamma(k)} \int_0^{(x/\eta)^m} t^{k-1} \exp(-t) dt \quad (x > 0, m > 0, \eta > 0, k > 0)$$

で与えられる。一般化ガンマ分布は尺度母数 η と二つの形の母数 m と k を持つ。特に一般化ガンマ分布の形の母数 k は分布族の違いを示す母数で、 $k=1$ の時にはワイブル分布を示し、 $k \rightarrow \infty$ の時には対数正規分布に近いことを示す。

2. データ変換による接近法

前節では、尺度母数と形の母数を持つ理論的な分布族を取り上げてERVデータに適合させることを考えた。この節では接近法を変えて、データ変換を行ってデータに適合させる立場、特に形の母数としてベキ母数を用いるベキ正規変換によってデータに適合させることを考える。ベキ母数の値の動きがワイブル分布や対数正規分布を含むベキ正規分布族として、新ベキ正規分布族(Isogai [3])を採用する。

新ベキ正規分布の分布関数 $F_{NP}(x)$ は正規分布の分布関数 Φ を用いて

(4)新ベキ正規分布

$$F_{NP}(x) = \Phi \left(\frac{1}{\delta} \left\{ \left(\frac{x}{\eta} \right)^{\frac{1}{\delta}} - \exp(-\delta^2) \right\} \right) \quad \left(-\frac{1}{3} < \delta < \frac{1}{3}, \eta > 0 \right)$$

で定義される。この分布は $\delta \rightarrow 0$ のとき対数正規分布に近づく。また $\delta \rightarrow 1/3$ のときにはワイブル分布

$$F_{NP}(x) = 1 - \exp \left(- \left(\frac{x}{\eta} \right)^{\frac{1}{\delta}} \right)$$

に、そして $\delta \rightarrow 1/3$ のときには、タイプ2の極値分布 (Johnson & Kotz [6]参照)

$$F_{NP}(x) = \exp \left(- \left(\frac{x}{\eta} \right)^{-\frac{1}{\delta}} \right)$$

に近づくことも判っている (磯貝 [4])。

δ の推定値の動く範囲についてであるが、実際に新ベキ正規分布を用いて様々なデータを解析してみると、最尤法を用いた場合に必ずしも厳密に $|\delta| < 1/3$ にならず、僅かばかり逸脱していることも多い。確認のために簡単な数値実験を行ってみたが、 δ の真値が境界に近い時にその傾向が顕著に現れる。従って、 δ の推定値は必ずしも母数空間に入るとは限らず、過剰推定や過小推定の状況が起こる。 δ の推定値が存在範囲を逸脱したとき、その最尤解が妥当かどうかは残差プロットなどで適合を評価するしか方法はない。そして、適合が妥当であると判断されたときには、 $\hat{\delta} \approx 1/3$ の場合に $\delta \rightarrow 1/3$ 、 $\hat{\delta} \approx -1/3$ の場合には $\delta \rightarrow -1/3$ と考えることにする。それでも不適合を引き起こすような大きな逸脱に対する対策としては、推定の範囲を限るしかなく、ここでは δ を探索する範囲として $|\delta| < 0.4$ でERVデータの解析を行う。

3. 尤度関数の構成

ERVが閾値 c よりも小さな値の時には自動的に0と記録されてしまうため、ERVの分布について、0が本当に出現しているのかどうか判らない状況になっている。そのため以下では二つの状況下 (内面塗膜が正常に塗布されてい

でもERVが必ずしも0にならない場合と内面塗膜が正常に塗布されていればERVが必ず0になる場合) について考察する. 前者の状況で取り扱う分布を打ち切りモデル, 後者の状況で取り扱う分布を混合モデルと呼ぶ. なお, ERVが正の値しか出現しないデータについては, 混合モデルと打ち切りモデルの区別がなくなるので, 打ち切りモデルの特殊な場合として取り扱う.

1. 打ち切りモデル

内面塗膜が正常に塗布されていてもERVが正になり得ると仮定する. この状況でERVが0になるのは, 計測器の閾値 c より下の値の時であると考え.

さて, 大きさ n の標本 $x_1, x_2, \dots, x_n$ が与えられたとして, 簡単のために最初の n_0 個の x が0, 残りの n_+

$(=n-n_0)$ 個の x が正であったとする. $x_1, x_2, \dots, x_n$ の同時密度から作られる尤度関数 L は,

$$L = [F_\theta(c)]^{n_0} \prod_{i=n_0+1}^n f_\theta(x_i)$$

で与えられる. θ の最尤推定量 $\hat{\theta}$ を

$$\frac{\partial \log L}{\partial \theta} = n_0 \frac{\partial}{\partial \theta} \log F_\theta(c) + \sum_{i=n_0+1}^n \frac{\partial}{\partial \theta} \log f_\theta(x_i) = 0$$

を解いて求める. ここで, $n_0 = 0$ の場合がERVデータがすべて正の状況に対応している.

2. 混合モデル

続いて内面塗膜が正常に塗布されているときには内面に欠陥なしと考えERVが0になり, 内面塗布に欠陥ある場合にはERVが正になると仮定する.

欠陥の有無を示す確率変数 Z を導入すると,

$$Z = 0 \Leftrightarrow \text{内面塗膜に欠陥無し} \Rightarrow ERV = 0$$

$$Z = 1 \Leftrightarrow \text{内面塗膜に欠陥有り} \Rightarrow ERV > 0$$

と書くことができる. Z の周辺分布については

$$\Pr(Z = 1) = p \quad (> 0), \Pr(Z = 0) = 1 - p \quad (> 0)$$

とする. さらに確率変数 X で表したERVについて, 欠陥がある場合の条件付き密度関数を $f_\theta(x)$ で, その分布

関数を $F_\theta(x)$ で表す. ただし θ は未知母数とする.

ここで, ERVが0と判定される上限の値である閾値 c を用いて, 実際に観測されるERVを確率変数 Y で表し, X との事象の関係を調べると

$$\{Z = 0\} \cup \{0 < X < c\} \Rightarrow \{Y = 0\}$$

$$\{X \geq c\} \Rightarrow \{Y = X\}$$

となっている.

さて, 以上の状況の下で, 大きさ n の標本 $y_1, y_2, \dots, y_n$ が与えられたとして, 簡単のために最初の n_0 個の y が 0, 残りの n_+ ($=n-n_0$) 個の y が正であったとする. このとき, $y_1, y_2, \dots, y_n$ の同時密度関数から作られる尤度関数 $\tilde{L}^*$ は

$$\begin{aligned}\tilde{L}^* &= (1-p^*)^{n_0} \prod_{i=n_0+1}^{n_0+n_+} [p^* f_\theta^*(y_i)] \\ &= (1-p^*)^{n_0} (p^*)^{n_+} \prod_{i=n_0+1}^n f_\theta^*(y_i)\end{aligned}$$

で与えられる(Isogai et al. [5]). ただし

$$\begin{aligned}p^* &= p(1-F_\theta(c)) \\ f_\theta^*(y) &= \frac{f_\theta(y)}{1-F_\theta(c)} \quad (y \geq c)\end{aligned}$$

と置いた. ここで, $n_0 > 0$ および $p=1$ のときには, 尤度関数 $\tilde{L}^*$ が打ち切りモデルの尤度関数 L と一致することとを注意しておく. また, $n_0 = 0$ の時には, $p=1$ とする. このときにも, 尤度関数 $\tilde{L}^*$ が打ち切りモデルの尤度関数 L と一致する. データセットの E (高) は, $n_0 = 0$ の場合である.

母数 p と θ の最尤推定量を求めるための尤度方程式を導くと

$$p(1-F_\theta(c)) = \frac{n_+}{n}$$

$$\sum_{i=n_0+1}^n \frac{\partial}{\partial \theta} \log f_\theta^*(y_i) = 0$$

なる同時連立方程式を解けばよいことが判る.

4. 分布の適合性の検討

2.2節で述べた三つの理論分布 Ratio 分布, Product 分布, 一般化ガンマ分布, および変数変換族の新ベキ正規分布を含む 4 つの分布について, 打ち切りモデルと混合モデルを用いて, 2.1 節のデータセットに当てはめた結果を表 3, 表 4 に示す. 表の欄には負の対数尤度の最小値と母数の推定値を示している.

1. 分布の適合性の比較

最初に打ち切りモデルと混合モデルの比較を行う。表3と表4の負の対数尤度の値を比較すると、データセット A（低）において、打ち切りモデルと混合モデルの差が明確に現れている。混合モデルの適合度が明らかに良い。

そこでは $\hat{p}$ の値がほぼ0.48となり、 $ERV=0$ となる場合の確率が約0.5であることを示している。一方、A（高）を

除く残りのデータセットについては、 $\hat{p}=1$ であり、打ち切りモデルと混合モデルの差は無さそうに見える。大概

の場合には、打ち切りモデルで済みそうであるが、基本的には混合モデルで取り扱う必要があると考えられる。

表3 打ち切りモデルでの4つの分布による解析結果
(各欄の数値は負の対数尤度の最小値と母数の推定値)

データセット	Ratio分布	Product分布	一般化ガンマ分布	新ベキ正規分布
A(高)	181.5 $\hat{\beta}=5.3e-9$ $\hat{\varphi}=1.1e-7$	119.2 $\hat{\beta}=0.1056$ $\hat{\varphi}=0.005$	116.2 $\hat{\eta}=5.87$ $\hat{m}=2.94$ $\hat{k}=0.0041$	119.1 $\hat{\eta}=5.4e-5$ $\hat{\sigma}=7.1333$ $\hat{\delta}=0.4$
A(低)	1601.9 $\hat{\beta}=2.9e-6$ $\hat{\varphi}=3.6e-7$	1651.7 $\hat{\beta}=16.26$ $\hat{\varphi}=0.006$	1369.4 $\hat{\eta}=66.8$ $\hat{m}=2.87$ $\hat{k}=0.04$	1456.9 $\hat{\eta}=1.522$ $\hat{\sigma}=4.200$ $\hat{\delta}=0.4$
B(高)	195.5 $\hat{\beta}=0.0563$ $\hat{\varphi}=0.71$	192.9 $\hat{\beta}=0.1394$ $\hat{\varphi}=0.81$	191.1 $\hat{\eta}=5.86e-22$ $\hat{m}=0.091$ $\hat{k}=69.4$	184.6 $\hat{\eta}=0.0593$ $\hat{\sigma}=0.7750$ $\hat{\delta}=-0.2680$
B(低)	1164.8 $\hat{\beta}=0.8512$ $\hat{\varphi}=0.24$	1173.3 $\hat{\beta}=4.616$ $\hat{\varphi}=0.38$	1166.9 $\hat{\eta}=0.033$ $\hat{m}=0.31$ $\hat{k}=3.59$	1167.5 $\hat{\eta}=1.8430$ $\hat{\sigma}=1.7025$ $\hat{\delta}=0.1523$
E(高)	-143.7 $\hat{\beta}=0.2657$ $\hat{\varphi}=3.50$	-146.5 $\hat{\beta}=0.3434$ $\hat{\varphi}=3.29$	-281.2 $\hat{\eta}=2.88e-23$ $\hat{m}=0.094$ $\hat{k}=115.1$	-303.8 $\hat{\eta}=0.1734$ $\hat{\sigma}=0.9575$ $\hat{\delta}=-0.1013$
E(低)	312.1 $\hat{\beta}=0.0664$ $\hat{\varphi}=0.073$	323.7 $\hat{\beta}=1.412$ $\hat{\varphi}=0.073$	307.3 $\hat{\eta}=7.89e-12$ $\hat{m}=0.11$ $\hat{k}=14.9$	306.8 $\hat{\eta}=0.1630$ $\hat{\sigma}=2.3034$ $\hat{\delta}=0.0143$

注： $e \pm q$ は $10^{\pm q}$ を示す。

表4 混合モデルでの4つの分布による解析結果

(各欄の数値は負の対数尤度の最小値と母数の推定値)

データセット	Ratio 分布	Product 分布	一般化ガンマ分布	新ベキ正規分布
A(高)	118.2 $\hat{\beta} = 1.16e-7$ $\hat{\varphi} = 1.26e-7$ $\hat{p} = 0.073$	119.2 $\hat{\beta} = 0.0968$ $\hat{\varphi} = 0.0064$ $\hat{p} = 0.921$	116.2 $\hat{\eta} = 5.874$ $\hat{m} = 2.945$ $\hat{k} = 0.0041$ $\hat{p} = 1$	118.2 $\hat{\eta} = 0.316$ $\hat{\sigma} = 2.42$ $\hat{\delta} = 0.4$ $\hat{p} = 0.087$
A(低)	1277.4 $\hat{\beta} = 17.81$ $\hat{\varphi} = 1.09e+8$ $\hat{p} = 0.482$	1277.4 $\hat{\beta} = 17.81$ $\hat{\varphi} = 5.71e+7$ $\hat{p} = 0.482$	1268.3 $\hat{\eta} = 21.33$ $\hat{m} = 1.340$ $\hat{k} = 0.892$ $\hat{p} = 0.480$	1262.4 $\hat{\eta} = 19.26$ $\hat{\sigma} = 0.771$ $\hat{\delta} = 0.352$ $\hat{p} = 0.485$
B(高)	195.5 $\hat{\beta} = 0.0557$ $\hat{\varphi} = 0.7161$ $\hat{p} = 1$	192.9 $\hat{\beta} = 0.1374$ $\hat{\varphi} = 0.8184$ $\hat{p} = 1$	191.4 $\hat{\eta} = 1.98e-16$ $\hat{m} = 0.1133$ $\hat{k} = 44.19$ $\hat{p} = 1$	184.6 $\hat{\eta} = 0.0593$ $\hat{\sigma} = 0.774$ $\hat{\delta} = -0.268$ $\hat{p} = 1$
B(低)	1164.8 $\hat{\beta} = 0.8439$ $\hat{\varphi} = 0.2419$ $\hat{p} = 1$	1173.3 $\hat{\beta} = 4.524$ $\hat{\varphi} = 0.3855$ $\hat{p} = 1$	1166.9 $\hat{\eta} = 0.0326$ $\hat{m} = 0.3127$ $\hat{k} = 3.604$ $\hat{p} = 1$	1167.5 $\hat{\eta} = 1.858$ $\hat{\sigma} = 1.700$ $\hat{\delta} = 0.154$ $\hat{p} = 1$
E(高)	-143.7 $\hat{\beta} = 0.2657$ $\hat{\varphi} = 3.50$ ($\hat{p} = 1$)	-146.5 $\hat{\beta} = 0.3434$ $\hat{\varphi} = 3.29$ ($\hat{p} = 1$)	-281.2 $\hat{\eta} = 2.88e-23$ $\hat{m} = 0.094$ $\hat{k} = 115.1$ ($\hat{p} = 1$)	-303.8 $\hat{\eta} = 0.1734$ $\hat{\sigma} = 0.9575$ $\hat{\delta} = -0.1013$ ($\hat{p} = 1$)
E(低)	312.1 $\hat{\beta} = 0.0441$ $\hat{\varphi} = 0.0475$ $\hat{p} = 1$	323.7 $\hat{\beta} = 1.439$ $\hat{\varphi} = 0.0609$ $\hat{p} = 1$	307.0 $\hat{\eta} = 1.21e-18$ $\hat{m} = 0.0833$ $\hat{k} = 27.14$ $\hat{p} = 1$	306.8 $\hat{\eta} = 0.1623$ $\hat{\sigma} = 2.303$ $\hat{\delta} = 0.013$ $\hat{p} = 1$

注1 : $e \pm q$ は 10^q を示す.注2 : E(高)の場合, $p=1$ を推定する必要がないので, ($\cdot$)をつけて示している.

また、表4でのA（高）の場合について、Ratioモデルでは打ち切りモデルよりも混合モデルの方に適合性の顕著な改善が見られ、 $\hat{p} = 0.073$ という値もERV=0の値がすこぶる多いというデータの分布状況を良く反映している。

一方、Product分布と一般化ガンマ分布については、ほぼ $\hat{p} = 1$ であって、打ち切りモデルと混合モデルの差がなく、A（高）のデータの分布状況を反映しているとは言い難い。新ベキ正規分布については、混合モデルと打ち切りモデルの適合については尤度の値に差はないが、 $\hat{p} = 0.087$ という値が得られデータの分布状況を良く反映している。

最後に、四つの分布の適合性の比較であるが、表4からはE（高）以外の場合には、四つの分布の適合性にそれほど大きな相違はないと思われる。しかし、表3（表4）のE（高）の結果からは、新ベキ正規分布が最も適合性が良いように見える。

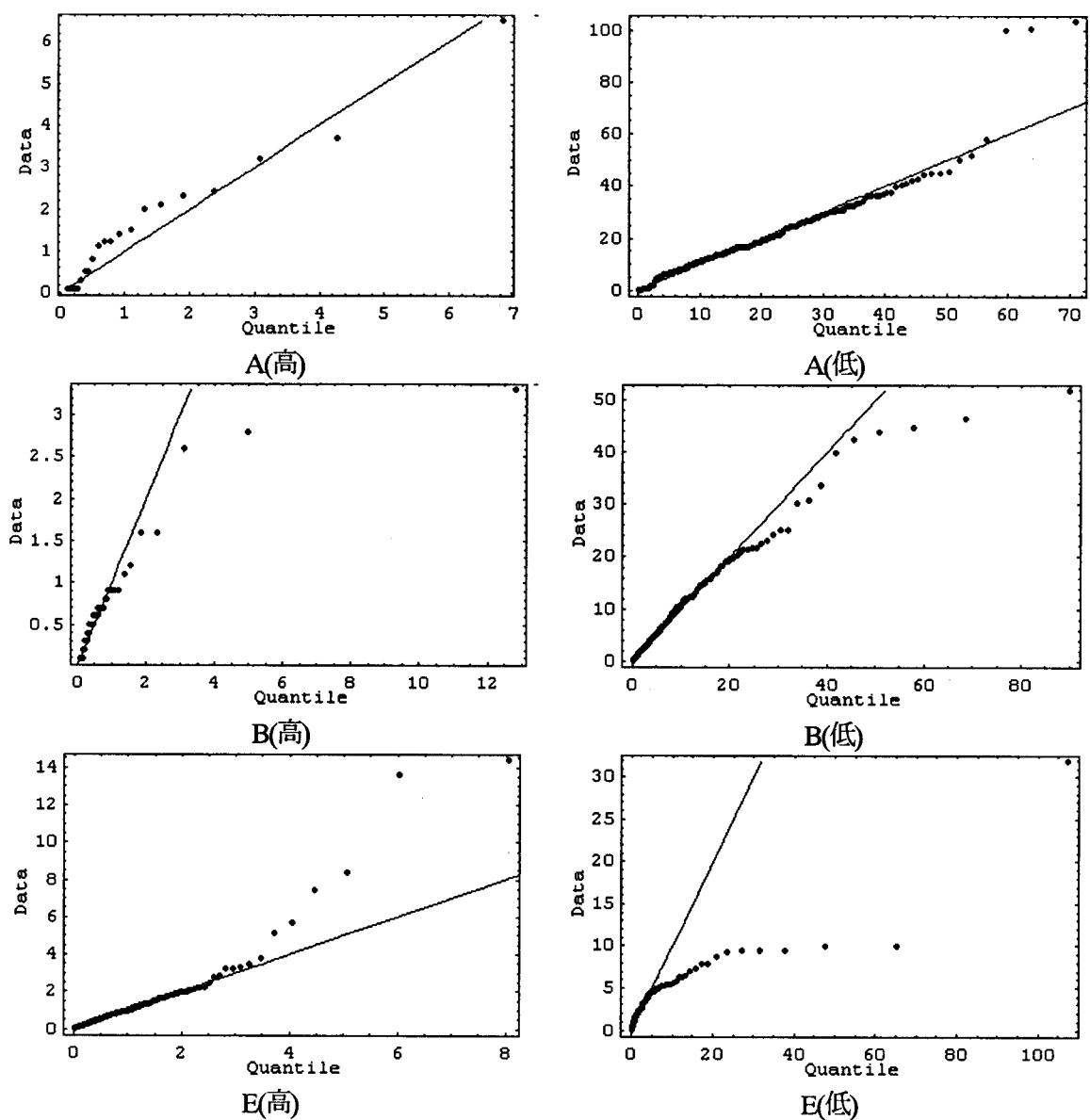


図2 混合モデルにおける新ベキ正規分布

2. 確率プロットによる適合性の検討

引き続きQ-Qプロットを用いて、新ベキ正規分布についてデータへの適合性を調べた結果を示す。混合モデルに対するQ-Qプロットの打点法は、正であるERVデータの順序統計量を $y_{(1)}y_{(2)}\cdots y_{(n_*)}$ としたとき、 $F_{\hat{\theta}}^*(y) = i/(n_*+1)$ ($i=1,2,\dots,n_*$) を解いた $\tilde{y}_{(i)}$ を用いて、点 $(\tilde{y}_{(i)}, y_{(i)})$ を打点する。ここで $\hat{\theta}$ はデータから求めた θ の推定値である。図2に混合モデルに関する新ベキ正規分布の結果を示す。

図2から、E(低)を除いて、新ベキ正規分布が良く適合していることが分かる。そこで、残りの三つの分布に関して、E(低)に対するQ-Qプロットを図3に与える。図3から、Product分布と一般化ガンマ分布は新ベキ正規分布の結果と差は見られないが、Ratio分布がデータセットE(低)の全体の特徴を良く捉えていることが分かる。E(低)の分布のより詳細な調査については、もう少し情報が必要である。

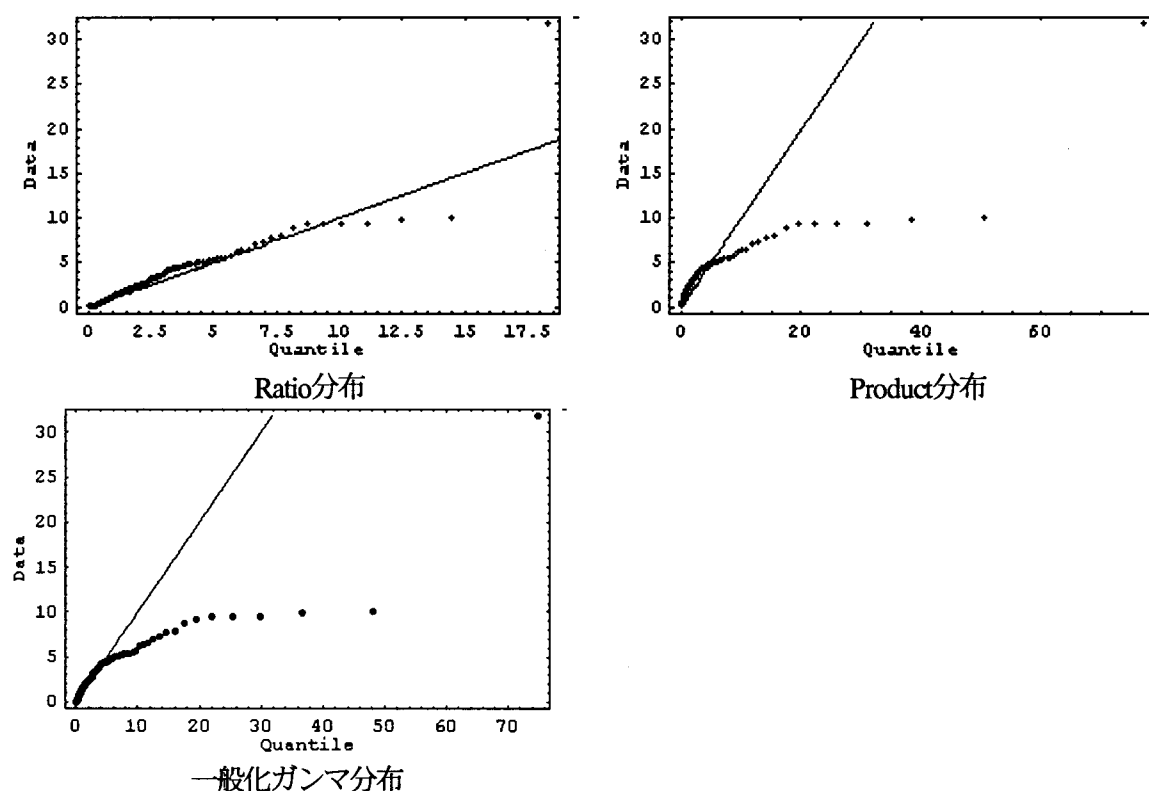


図3 データセットE(低)に対するQ-Qプロット

3. 幾つかの注意点

- (a) E(低)の分布について、米国の研究者から対数正規分布であるとの指摘を受けていた。表3と表4の結果から、一般化ガンマ分布では δ の値がほぼ15以上であり、新ベキ正規分布では δ の値が0に近いと考えられる。そのことにより、我々の結果もそれを再確認したことになる。
- (b) 一般化ガンマ分布の解析結果の中に、尺度母数 η の値がほとんど0と見なされるものがある。これは実際の適用に当たって、大きな障害となる。また、数値計算で収束の遅いことも今回の解析の中で確認されている。
- (c) 新ベキ正規分布の形の母数 δ の動きであるが、表3および表4で、E(高)とB(高)の δ の符号が同じであり、E(低)、B(低)およびA(低)の δ の符号が同じであることから、腐食性の相対的に高い飲料用と低い飲料用のもののERVデータの分布の型が異なることが予想される。しかし、A(高)については、他の腐食性の相対的に高い飲料用

のERVの分布とは δ の符号が逆である。そこで、表 3 の打ち切りモデルで、 δ の値を動かしてA(高)の適合度を調べると、 $\delta = -0.1$ の時に、負の対数尤度の最小値=119.8, $\hat{\eta} = 1.1e-4$, $\hat{\sigma} = 3.4666$ となり、適合度がほとんど変わらない。 δ の値の符号が変わっても適合度が変わらない理由は、A(高)のデータの中で正のERVを持つものが、 $n=500$ の中で23個しかなく推定精度が悪いためだと思われる。

表4の混合モデルでも同様の考え方で、 δ の値を動かしてA(高)の適合度を調べた。 $\delta = -0.1$ の時に、負の対数尤度の最小値=119.4, $\hat{\eta} = 0.0825$, $\hat{\sigma} = 2.1532$, $\hat{p} = 0.091$ となり、適合度がほとんど変わらないことが分かる。腐食性の相対的に高い飲料用と低い飲料用のもののERVの分布の型が異なる可能性が再び示唆された。

5. まとめ

ERVデータの分布を調べるために、打ち切りモデルと混合モデルという2種類の尤度関数を構成し、4つの確率分布(Product分布, Ratio分布, 一般化ガンマ分布, 新ベキ正規分布)に関して、それらの適合性を検討した。得られた結論は以下のようにまとめられる。

- (1) 尤度関数としては混合モデルが適切である。
- (2) 確率分布としては、データセットE(高)を除いて、4つの分布の適合度にそれほど大きな差はない。データセットE(高)を含めると新ベキ正規分布が最も適合性が高く、次いで一般化ガンマ分布, Ratio分布という順になっている。
- (3) この調査でも一般化ガンマ分布の欠点である数値計算上の取り扱いにくさが現れ、ERVデータへの適用には問題があると思われる。
- (4) 新ベキ正規分布を用いることで、ERVデータの分布について、腐食性の相対的に高い飲料用と低い飲料用とでは分布の型が異なっている可能性が指摘されている。残りの3つの分布では分布の型の違いを見分けることは難しい。

以上のことから、ERVデータに対する分布としては、取り扱いやすさと適用性の広さの点から、変数変換を用いる新ベキ正規分布が勧められる。ただし、これはあくまで実用的な観点からの結論である。

6. 参考文献

- [1] Isogai, T. & Uchida, H. (1989): "Some measurement models related to an Inverse Gaussian Distribution", Reports of Statistical Application Research, JUSE, 36, 1-11.
- [2] Isogai, T. (1993): "Note on some measurement models related to an Inverse Gaussian Distribution", Reports of Statistical Application Research, JUSE, 40, 1-12.
- [3] Isogai, T. (1999): "Power transformation of the F distribution and a power normal family", Journal of Applied Statistics, 26, 355-371.
- [4] 磯貝恭史 (2001): "F分布のベキ変換公式とそれから導かれる新ベキ正規分布族の応用", Journal of the Japanese Society for Quality Control, 31, 89-104.
- [5] Isogai, T., Uchida, H., Miyama, S. & Nishiyama, S. (2006): "Statistical Modeling of Enamel Rater Value Data", (submitted).
- [6] Johnson, N. L. & Kotz, S. (1970): "Continuous Univariate Distributions-1", John Wiley & Sons.
- [7] 西山貞雄 (1988): "アルミニウム缶の腐食防食", Boshoku Gijutsu, 37, 578-579.
- [8] 西山貞雄・栗原精二・当摩 建 (1994): "オールアルミニウム缶での微小塗膜欠陥部の検出と面積測定に関する研究", 材料と環境, 43, 11-17.