極値統計について

エムエス・ラボ　　小　山　信　次

１．統計の基礎的事項

１．１　度数分布表とヒストグラム

腐食の場合を例に取ると、サンプリングした腐食データを腐食量を範囲に分け、範囲に入る度数を調べ、度数分布表に表す。表 1 では 0.1mm 間隔に分け、度数 A を調べている。この場合、減肉量(mm)が確率変数 x になり、確率密度 f は、各度数 A をサンプリング データ総数 N に 1 を加えた値で割った値、A/(N+1)になる。これを平均ランク法と言う。表 1 の場合は サンプリング データ総数 N=100 である。

表 1　度数分布表

確率変数x mm	度数 A 個数	確率密度f A/(N+1)	累積分布 F = Σf	備考
0.4-0.5	21	0.208	0.208	
0.5-0.6	50	0.495	0.703	F=0.208+0.498=0.703
0.6-0.7	2	0.020	0.723	F=0.208+0.498+0.020=0.723
0.7-0.8	4	0.040	0.763	・・・・・・・
0.8-0.9	5	0.050	0.813	・・・・・・・
0.9-1.0	6	0.059	0.872	・・・・・・・
1.0-1.1	3	0.030	0.902	・・・・・・・
1.1-1.2	8	0.079	0.981	・・・・・・・
1.2-1.3	0	0.000	0.981	・・・・・・・
1.3-1.4	1	0.010	0.991	F=0.208+・・・・・・・+0.010=0.991
(減肉量)	(N=100)	(N+1=101)		

ヒストグラムは 確率変数 x と度数 A の関係を棒グラフに表したものである。

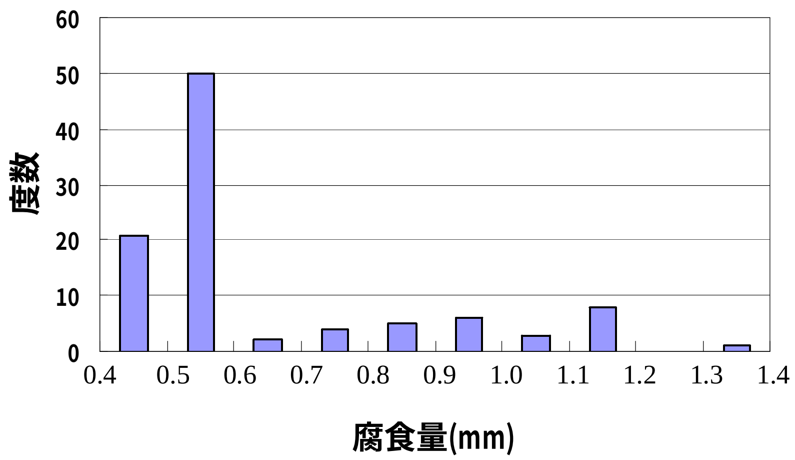


図 1　腐食データのヒストグラム

1.2　密度関数f(x)

度数を確率変数 x の関数として A(x)と定義すると 確率密度関数 f(x)は次の関係にある。関連する事項について次式に示す。N はサンプリング データ総数

$$f(x)=\frac{A(x)}{\int_{-\infty}^{+\infty}A(u)du} \qquad \int_{-\infty}^{-\infty}f(x)dx=1 \qquad f(x)\geqq 0 \qquad \int_{-\infty}^{+\infty}A(x)dx=N$$

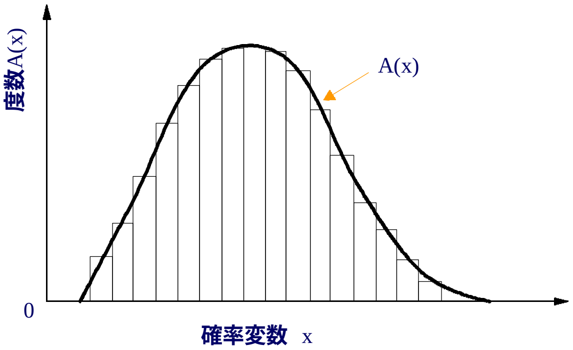


図 2　度数分布関数 A(x)

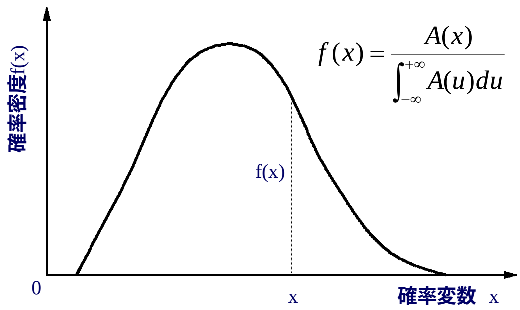


図 3 確率密度関数 f(x)

1.3　累積分布関数 F(x)

表 1 の累積分布を関数 F(x)とする。確率密度関数 f(x)との関係は次式のようになる。

$$F(x)=\int_{-\infty}^{x}f(u)du \qquad \frac{dF(x)}{dx}=f(x) \qquad \int_{-\infty}^{\infty}f(u)du=1$$

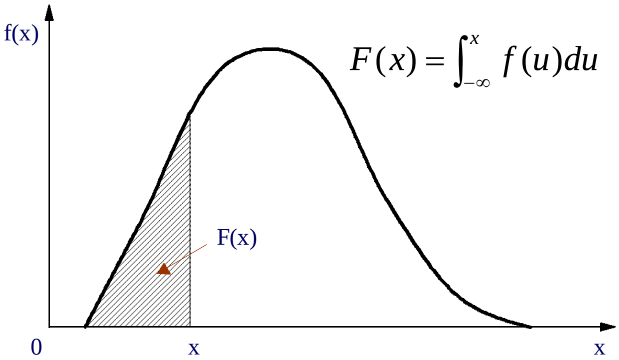


図 4　累積分布関数 F(x)

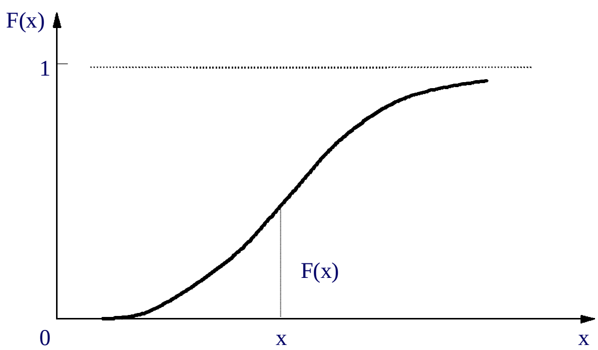


図 5　累積分布関数 F(x)

2.　極値統計の理論とその方法

●均一腐食　→　平均値問題

●孔食・すきま腐食・応力腐食割れ等の局部腐食　→　極値問題

多数の局部腐食のうち、最も深く進行し、材料の肉厚を貫通するまでの時間がその材料の寿命。石油タンク底板の寿命を考える場合に、局部腐食の深さに関して、その最大値が問題となる。

2.1　腐食現象の極値分布について

ある石油タンクの底板の裏面腐食の最大値を推定する。一定面積をもつ領域を n 箇所選び、それぞれの領域から得られるデータからタンク底板全面積における最大腐食量を推定・

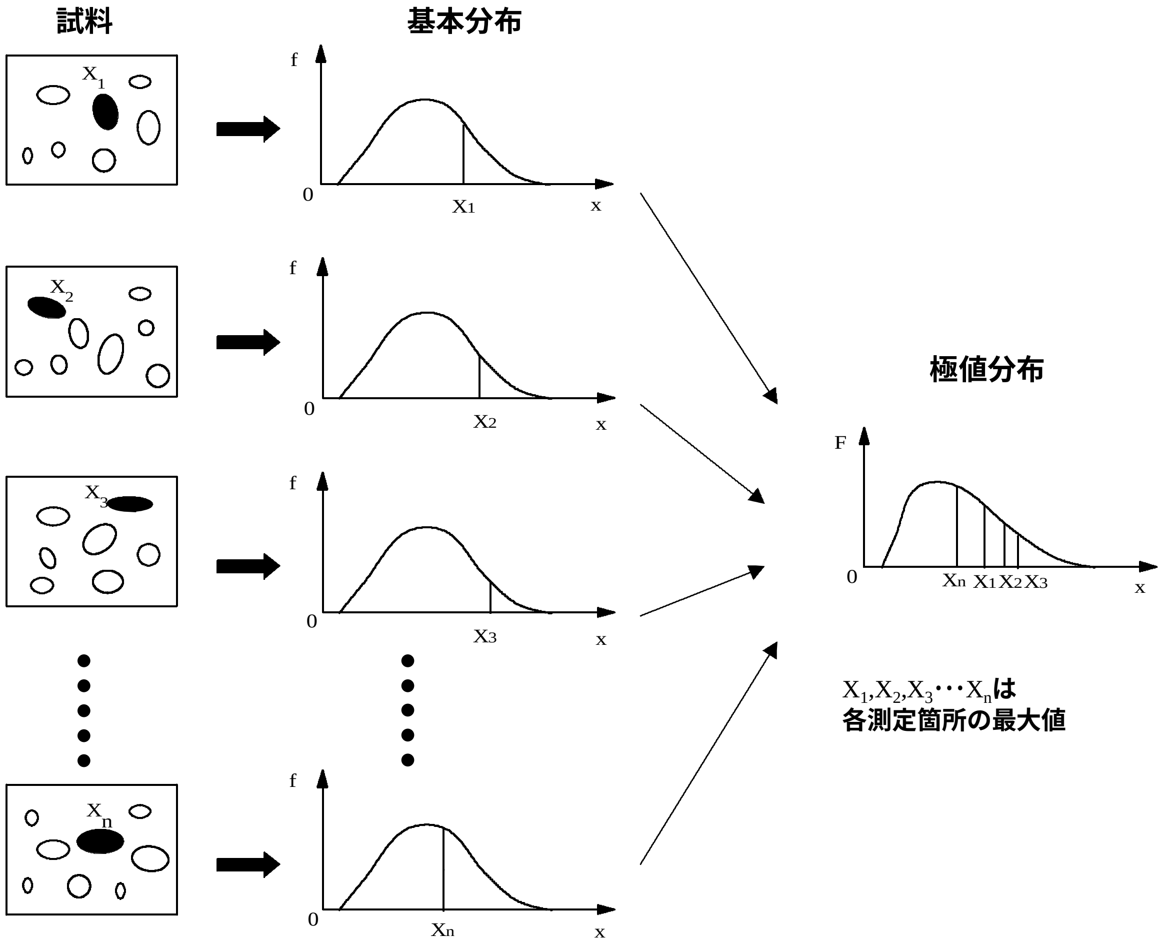


図 6　最大値の度数分布

n 箇所の測定値の最大値をそれぞれ、X_1, X_2, ･･････X_i ･･･X_n　とする。最大値 X_1, X_2, ･･････X_i ･･･X_n の分布が極値分布である。表 2 に n=23 の場合の最大値の度数分布表を示した。

最大値の度数分布から以下の値を計算する。

(1).累積度数F_iの計算

最大値の分布を検討する場合、度数（f_i） を小さい方から累積し、F_i を計算する。

(2).累積分布$F_i/(n+1)$の計算と二重指数確率紙へのプロット

累積度数（F_i）を試料個数nに１を加えた値（n+1）で割って累積分布（$F_i/(n+1)$）を算出する。nはデータ数である。平均ランク法という。結果を表.2 に示した。これを二重指数確率紙に、x軸上に腐食量を目盛り、各クラスの上限値をx値とし、$F_i/(n+1)$を$F_I(x)$値としてプロットする。表 2 の場合、n=23 である。１カ所当たりのサンプリングデータ数は 100 である。

表２　最大値の度数分布

クラス	腐食量 x（mm）	度数 fi	累積度数 Fi	累積分布 Fi／（n＋1）
1	0.95－1.00	1	1	0.042
2	1.00－1.05	3	4	0.167
3	1.05－1.10	1	5	0.208
4	1.10－1.15	1	6	0.250
5	1.15－1.20	1	7	0.292
6	1.20－1.25	0	7	0.292
7	1.25－1.30	1	8	0.333
8	1.30－1.35	2	10	0.417
9	1.35－1.40	4	14	0.583
10	1.40－1.45	0	14	0.583
11	1.45－1.50	1	15	0.625
12	1.50－1.55	2	17	0.708
13	1.55－1.60	0	17	0.708
14	1.60－1.65	3	20	0.833
15	1.65－1.70	1	21	0.875
16	1.70－1.75	0	21	0.875
17	1.75－1.80	0	21	0.875
18	1.80－1.85	1	22	0.917
19	1.85－1.90	0	22	0.917
20	1.90－1.95	0	22	0.917
21	1.95－2.00	0	22	0.917
22	2.00－2.05	0	22	0.917
23	2.05－2.10	0	22	0.917
24	2.10－2.15	0	22	0.917
25	2.15－2.20	1	23	0.958

横軸　　　　縦軸

２．２　二重指数分布の適用の判断

二重指数分布は、**Gumbel** 分布とも呼ばれている。これには，最大値分布と最小値分布がある。局部腐食による損傷おいて一番問題となるのは、その最大侵食量、それによる装置の最小寿命の予測であることに対応している。

最大値分布 $F_{I}(x)$、最小値分布 $F_{-I}(x)$ の式は次式によって与えられる。

$$F_{I}(x)=\exp[-\exp\{-(x-\lambda)/\alpha\}] \qquad -\infty<x<\infty$$

$$F_{-I}(x)=1-\exp[-\exp\{-(x-\lambda)/\alpha\}] \qquad -\infty<x<\infty$$

λ:位置パラメータ[=最頻値]　　α:尺度パラメータ

それぞれの密度関数は、

$$f_I(x) = (1/\alpha)\cdot\exp[-(\alpha-\lambda)/\alpha - \exp\{-(x-\lambda)/\alpha\}]$$
$$f_{-I}(x) = (1/\alpha)\cdot\exp[(\alpha-\lambda)/\alpha - \exp\{(x-\lambda)/\alpha\}]$$

で与えられる。

ここで，α、λは位置パラメータ，尺度パラメータと呼ばれ，α＞0である。密度関数が導かれる過程において、xが大きい領域で基本分布が指数分布となることが仮定されている。極値解析を行う場合、基本分布が不明であっても手続き上不都合はないのであらかじめチェックする必要はない。ただし、最大値のデータが推定された二重指数分布に適合しているかどうかは検定によって確かめることになる。

２.３　二重指数確率紙を利用したタンク底板の寿命推定

最大値分布$F_I(x)$の両辺の自然対数を２回とる

$$F_I(x) = \exp[-\exp\{-(x-\lambda)/\alpha\}]$$

λ :位置パラメータ[=最頻値]

α :尺度パラメータ

$$-\ell n\{-\ell n F_I(x)\} = (x-\lambda)/\alpha$$

ここで，　$y = -\ell n\{-\ell n F_I(x)\}$

とおくと

$$y = (1/\alpha)x - (\lambda/\alpha)$$

y は二重指数分布における規準化変数と定義する。

$F_I(x) = \exp(-e^{-y})$　　$-\infty<x<\infty$

$F_{-I}(x) = 1 - \exp(-e^{-y})$　　$-\infty<x<\infty$

●二重指数確率紙

二重指数確率紙は、直交座標系の１軸に

$$y = -\ell n\{-\ell n F_I(x)\}$$

を目盛り，他の一軸に確率変数xを目盛った方眼紙である。縦軸に$F_I(x)$、横軸にxにとり、データをプロットして，直線のまわりのばらつきが一定の範囲内にあれば二重指数分布に従っていると判断される。この直線からパラメータα、λを求める。

$y = -\ell n\{-\ell n F_I(x)\}$　の関係から、y = 0 とおくと　$\ell n\{-\ell n F_I(x)\} = 0$

$-\ell n F_I(x) = 1$　　$\therefore -\ell n F_I(x) = 1$　　$F_I(x) = e^{-1} = 0.368$

$$y = (1/\alpha)x - (\lambda/\alpha) = 0 \qquad \therefore\ x = \lambda$$

となるので、y = 0 、$F_I(x)$ = 0.368 を与える x の値がλとなる。

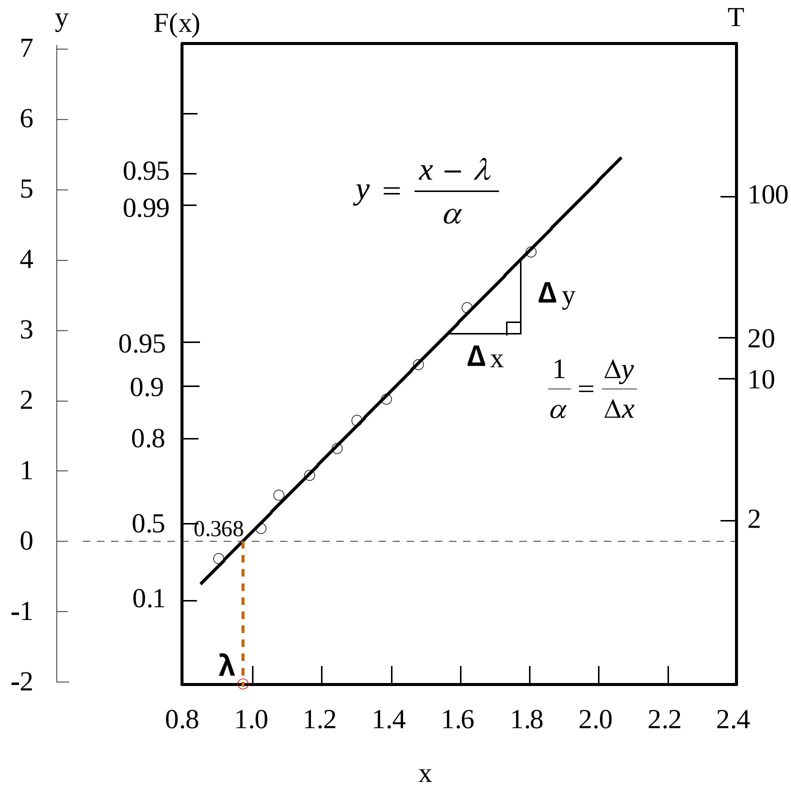


図７　二重指数確率紙　　プロットデータは説明用

一方、直線の傾きから

$$\frac{1}{\alpha} = \frac{\Delta y}{\Delta x} \qquad \therefore \alpha = \frac{\Delta x}{\Delta y}$$

のようにαを求めることができる。

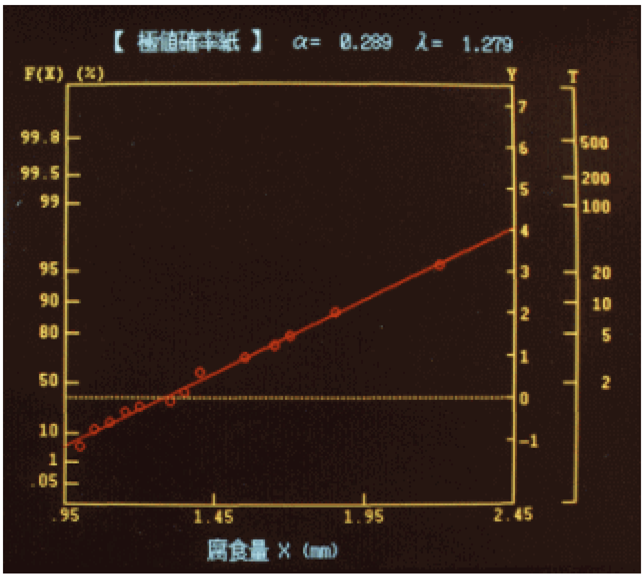


図８　コンピュータ処理した場合の確率紙でのデータ
直線は最小二乗法により決定している。α、λの値も自動的に計算

表２のデータの パラメータλ, αを求めた結果、y=3.4573 x − 4.426 が求められたので、データ例の場合は、α=0.289、λ=1.253 が得られた。分布関数$F_I(x)$は、データ例の場合は、

$$F_I(x) = \exp[-\exp\{-(x-1.253)/0.289\}]$$

となる。

２.４ 寿命予測の方法

１つのサンプル領域から，１個の最大値ｕが得られる場合に、ある値ａより大きな値が得られるまでには何カ所の領域について調べることになるのかを考えてみる。今，この数をＮとすると、その確率はＰ(Ｎ) は次のようになる。はじめから（Ｎ－１）箇所までのｕのそれぞれがａより小さい確率がＦ(ａ)であるから、(Ｎ－１）箇所すべてにおいて、ａより小さい確率は$F(a)^{N-1}$である。Ｎ番目の領域で、ｕがａより大である確率は｛１－Ｆ(ａ)｝である。従って、

$$P(N) = \{F(a)\}^{N-1} \cdot \{1-F(a)\} \qquad (4)$$

Ｎの期待値Ｔは再帰期間である。

Ｆ＜１であるから

$$T = 1 / \{1-F(a)\} \qquad (5)$$

となり、確率紙におけるＴ目盛りとＦ(ｘ)目盛りの関係は（1）式の関係がある。

次にＴを設定する。今考えている全底板を対象領域に考えると、次式が成り立ち、データを適用すると次式の結果となる。

【例１】

Ｔ＝（全低板面積）／（ サンプル領域１つの面積）

＝ $(\pi/4) \cdot D^2 / (10 \times 10)$

＝ 5,772,300

ただし，タンク直径 D=27.11m、測定領域は 10 mm × 10 mm である。

Ｔに対応するｘ（x_{max}とおく）を求めると、

$-\ell n\{-\ell n F_I(x)\} = (x-\lambda)/\alpha$ と $T = 1/\{1-F(a)\}$ 式から

$$x_{max} = \lambda + \alpha \cdot \ell n\{-\ell n(1-1/T)\}$$

となり，Ｔ＞＞１であるから,

$$x_{max} = \lambda + \alpha \cdot \ell n T$$

T = 577230, λ = 1.253, α=0.289 を上式に代入すると、x_{max} = 5.473 が得られる。

残存寿命をＺ[年] とすると、

Ｚ＝（残存板厚）／（腐食速度） ＝（ｔ$_o$－ｔ）／Ｙ ＝ 0.0025 年

ここに、初期板厚ｔ$_o$＝ 6.0 mm、現状の最小板厚（x_{max}）、ｔ= 5.473、タンクの使用年数Ｙ＝ 15 年である。

表 3 タンク仕様

供 試 材	T−13
試料採取年月日	S．51．12
タンク完成年月日	S．36．11
タンク使用年数	15
油 種	重 油
形 式	コーンルーフ
直径 (m)	27.11
高さ (m)	15.13
容量 (kl)	7900
底板材質	SS41
アニュラー板厚(mm)	8
底板板厚(mm)	6
側板板厚(mm)	6〜18

【例 2】 サンプリング領域の 500 倍の面積を有する試験片に生ずる最大孔食深さを推定する

500 倍の面積を有する試験片は試験片数 N=500 の小面積試験片をサンプリングした場合と同一であり、材質も同じであれば確率紙で同一直線上に乗る。F=N/(N+1)=500/501 の再帰期間 T を計算すると T=1/(1-F)=N+1=501, 501 500 , T N

点 A を通る横軸と縦軸に平行な直線を引くと B 点の x の値が求める 500 倍の面積を有する試験片に生ずる最大孔食深さと推定できる。

＊【例１】では数式からx_{max}を求めたが、【例 2】では確率紙から図式的にx_{max} を求めた。

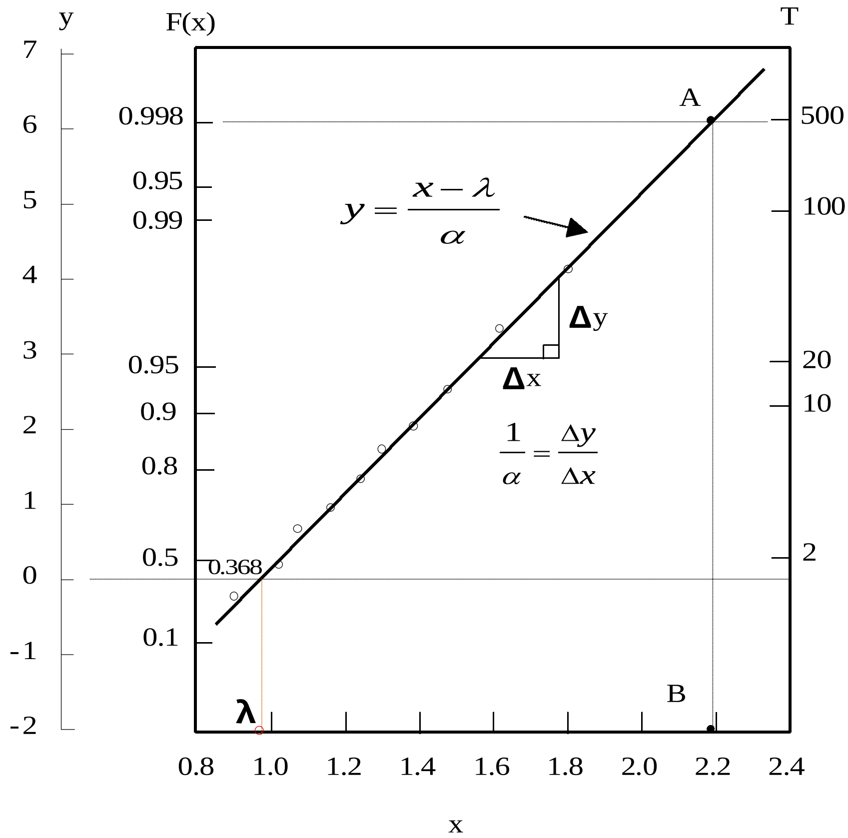


図 9

2.5　推定された確率分布の妥当性の検討

これまでに述べたような方法で、推定された二重指数確率分布が実測値を表しているものであるか否かを判定する必要がある。χ^2検定と K.S 検定法がある。ここでは、χ^2検定を得られた結果に対して適用する。

観測されたデータを k 個のクラスに分けて i 番目のクラスの下限を β_i,上限を β_{i+1}、また、このクラスにデータの数を Ni とし、全データ数を N とする。このとき、i 番目に含まれるデータの期待度数は次式で与えられる。

$$N \cdot P_i = N \cdot \int_{\beta_i}^{\beta_{i+1}} f(x)dx$$

このとき

$$\chi^2 = \sum \frac{(N_i - N \cdot p_i)^2}{N \cdot p_i}$$

は、自由度 k-C-1 の χ^2 分布に従うことが知られている。C は未知の分布のパラメータの数である。測定されたデータを使って当てはめようとする分布のパラメータを推定した場合、パラメータは未知であるとして扱われる。χ^2検定は求めた値 χ^2 が χ^2 分布表から求めた値より小さければ分布に適合すると考える。

χ²分布表から求める値は、今ある定数αを考えて

$$\alpha = p\left\{\chi^2 \le a\right\} = \int_a^{\infty} h_{k-1}(x)dx$$

を満たすαである。ここに、$h_{k-1}(x)$　は

$$h_k(x) = \begin{cases} 0 & for \quad x \le 0 \\ \dfrac{1}{2^{\frac{k}{2}\Gamma(\frac{k}{2})}} x^{\frac{k}{2}-1} \cdot e^{-\frac{k}{2}} & for \quad x > 0 \end{cases}$$

であり、Γ(x)はΓ関数である。

f(x)に、ある値aをとり、上式に代入すると、

$$\int_a^{\infty} f_n(x)dx = \alpha^*$$

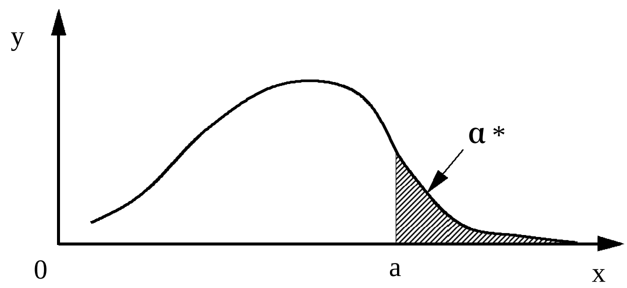


図 10

α*はχ²値すなわち、xがaよりも大きい確率を与える。このα*が十分小さく、（a≦x）とすると、そのようなx値が発生することはまれであり、N_iは実測値であるからxがaより大きくなってしまったのは、N•Pi　すなわち推定した理論値に問題があったわけである。すなわち、理論式が適切ではないと判定される。この判定内容はα*の値によって左右される。α*は危険率あるいは有意水準と呼ばれる。

【計算例】

石油タンク底板腐食量のサンプリングデータを表 4 のようにクラス分けし、腐食量の範囲に存在する実測値から、推定された二重指数分布が実測値を表しているかどうか判断するために、このデータの場合のχ²検定を行う。

i 番目のクラスに含まれるデータの期待度数は次式で与えられる。

$$N \cdot P_i = N \cdot \int_{\beta_1}^{\beta_2} f(x)dx$$

具体的に、Table.3 で、i=2 のとき、β_1 = 1.25、β_2 = 1.55　である。

F(x)と f(x)の関係、

$$\frac{dF(x)}{dx} = f(x)$$

から　$P_i = F(\beta_2) - F(\beta_1)$　となり、　$F_I(x)$は次式で与えられている。

$$F_I(x) = \exp[-\exp\{-(x-1.253)/0.289\}] \qquad \lambda = 1.253,\ \alpha = 0.289$$

データの期待度数は

$$N \cdot P_i = N\ \{F(\beta_2) - F(\beta_1)\} \qquad N=23$$

となる。

　同様に各クラスのχ^2値は　$(N_i - N \cdot P_i)^2 / N \cdot P_i$ で計算される。計算結果を表 4 に示した。

表 4

クラス	腐食量	実測値N_i	期待度数$N \cdot p_i$	$(N_i - N \cdot p_i)^2/NP_i$
1	0.95-1.25	7	7.047	0.000
2	1.25-1.55	9	7.708	0.185
3	1.55-1.85	5	4.181	0.160
4	1.85-2.15	0	1.728	1.228
5	2.15-2.45	2	0.647	2.832
計		23		4.405

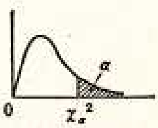
　ここで，上記の方法で推定された二重指数分布が実測値を表しているかどうか判断するために、このデータの場合のχ^2検定を行う。計算の結果、χ^2値は 4.405 となった。ここで、χ^2 分布表を参照することになるが、自由度kは、k=（クラス数－パラメータ数－１）である。クラス数 5，分布のパラメータ数はαとλの 2 つ。k=5-2-1 =2 となる。有意水準*を５%とすると、

$$\chi = 4.405,\quad a = 5.991$$

となり、$\chi < a$である。従って二重指数分布(3)式は５%の危険率で棄却されない、すなわち二重指数分布の仮定は正しいと言うことができる。

表 5　　χ^2 分布表

$P(\chi^2 \geqq \chi_\alpha^2)\alpha$



k \ α	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	$.0^3157$	$.0^3628$	$.0^2393$	.0158	.0642	.148	.455	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	.0201	.0404	.103	.211	.446	.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	.115	.185	.352	.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.345	16.266
4	.297	.429	.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.467
5	.554	.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.086	20.515

表 6　極値分布の種類 [1)]

	最大値分布	最小値分布
タイプI (二重指数分布)	$F_1(x)=\exp\left[-\exp\left(-\frac{x-\lambda}{\alpha}\right)\right]$ $f_1(x)=\frac{1}{\alpha}\exp\left[-\frac{x-\lambda}{\alpha}-\exp\left(-\frac{x-\lambda}{\alpha}\right)\right]$ $-\infty<x<\infty,-\infty<\lambda<\infty,\alpha>0$	$F_{-1}(x)=1-\exp\left[-\exp\left(\frac{x-\lambda}{\alpha}\right)\right]$ $f_{-1}(x)=\frac{1}{\alpha}\exp\left[\frac{x-\lambda}{\alpha}-\exp\left(\frac{x-\lambda}{\alpha}\right)\right]$ $-\infty<x<\infty,-\infty<\lambda<\infty,\alpha>0$
タイプII	$F_{\mathrm{II}}(x)=\exp\left[-\left(\frac{x-\gamma}{\eta}\right)^{-m}\right]$ $f_{\mathrm{II}}(x)=\frac{m}{\eta}\left(\frac{x-\gamma}{\eta}\right)^{-m-1}\exp\left[-\left(\frac{x-\gamma}{\eta}\right)^{-m}\right]$ $-\infty<\gamma\le x<\infty,\eta>0,m>0$	$F_{-\mathrm{II}}(x)=1-\exp\left[-\left(\frac{\gamma-x}{\eta}\right)^{-m}\right]$ $f_{-\mathrm{II}}(x)=\frac{m}{\eta}\left(\frac{\gamma-x}{\eta}\right)^{-m-1}\exp\left[-\left(\frac{\gamma-x}{\eta}\right)^{-m}\right]$ $-\infty<x\le\gamma<\infty,\eta>0,m>0$
タイプIII	$F_{\mathrm{III}}(x)=\exp\left[\left(-\frac{\gamma-x}{\eta}\right)^{m}\right]$ $f_{\mathrm{III}}(x)=\frac{m}{\eta}\left(\frac{\gamma-x}{\eta}\right)^{m-1}\exp\left[-\left(\frac{\gamma-x}{\eta}\right)^{m}\right]$ $-\infty<x\le\gamma<\infty,\eta>0,m>0$	$F_{-\mathrm{III}}(x)=1-\exp\left[-\left(\frac{x-\gamma}{\eta}\right)^{m}\right]$ $f_{-\mathrm{III}}(x)=\frac{m}{\eta}\left(\frac{x-\gamma}{\eta}\right)^{m-1}\exp\left[-\left(\frac{x-\gamma}{\eta}\right)^{m}\right]$ $-\infty<\gamma\le x<\infty,\eta>0,m>0$

表 7　腐食データの基本分布 [1)]

名 称	特 徴
Poisson分布 (Poisson distribution)	○離散分布 ○腐食現象がランダムに発生することを示す分布
指 数 分 布 (exponentialdistribution)	○腐食現象の故障率が常に一定であることを示す 分布
正 規 分 布 (normaldistribution)	○最も一般的に統計処理に用いられる分布
対数正規分布 (logtnormaldistribution)	○確率変数 (腐食データの特性値) の対数が物理 的意味をもつ場合の正規分布

表８　腐食データの統計処理に使用する分布[1]

分類	具体的名称	確率密度関数f(x), 累積分布関数F(x)	パラメータ
基本分布	Poisson分布	$f(x) = e^{-\alpha}\dfrac{\alpha^{x}}{x!}$	α
	指数分布	$f(x) = \dfrac{1}{\theta}\exp\left(-\dfrac{x}{\theta}\right)$　$F(x) = \int_0^x f(u)du = 1 - \exp\left(-\dfrac{x}{\theta}\right)$	θ
	正規分布	$f(x) = \dfrac{1}{\sqrt{2\pi}\sigma}\exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$　$F(x) = \int_{-\infty}^x f(u)du$	μ, σ
	対数正規分布	$f(x) = \dfrac{1}{\xi x\sqrt{2\pi}}\exp\left[-\dfrac{1}{2}\left(\dfrac{\ell nx - \ell n\delta}{\xi}\right)^2\right]$　$F(x) = \int_{-\infty}^x f(u)du$	δ, ξ
極値分布	二重指数分布	$f_1(x) = \dfrac{1}{\alpha}\exp\left[-\dfrac{x-\lambda}{\alpha} - \exp\left(-\dfrac{x-\lambda}{\alpha}\right)\right]$　$F_1(x) = \exp\left[-\exp\left(-\dfrac{x-\lambda}{\alpha}\right)\right]$	λ, α
	Weibull分布	$f_{-\mathrm{III}}(x) = \dfrac{m}{\eta}\left(\dfrac{x-\gamma}{\eta}\right)^{m-1}\exp\left[-\left(\dfrac{x-\gamma}{\eta}\right)^m\right]$　$F_{-\mathrm{III}}(x) = 1 - \exp\left[-\left(\dfrac{x-\gamma}{\eta}\right)^m\right]$	γ, η, m

参考文献

1）装置材料の寿命予測入門―極値統計の腐食への適用　、腐食防食協会

HP　http://MS-Laboratory.jp/