國澤清典　著　エントロピー・モデルの１因子情報路のケースの検証

2015.3.23 株式会社　応用数理研究所　佐々木俊久

１因子情報路は、國澤先生の本[註1]には下記のように記述されている。

「いま，n 銘柄 $A_1, A_2, \cdots, A_{n,}$ があり，ある特定の因子についてのそれぞれの
評価値を $l_1, l_2, \cdots, l_n$ とし，販売比率を $p_1, p_2, \cdots p_n$ とする。

Ⅰ　平均特性値 $\bar{l} = \sum_i l_i p_i$ をできるだけ小さくしたい。

Ⅱ　Ⅰと同時に，選択のあいまいさを表すエントロピー

$$H = -\sum_i p_i \log p_i$$ をできるだけ大きくしたい。」

この２つを満足するような販売比率の選択 $p_1, p_2, \cdots p_n$ を求める。Ⅰ，Ⅱの２つの条件が１因子情報路の特徴づけになっている。

この本では，評価値を互いに素な整数比にして，

$W^{-l_1} + W^{-l_2} + \cdots + W^{-l_n} = 1$　について正根 W_0 を求め，選択比率を

$p_i = W_0^{-l_i}$，$i = 1,2,\cdots n$

によって与えている。

この問題を、PC 上で、いくつかの方法で解いてみる。

(1) $H/\bar{l}$　正根を求めず、最大化問題として直接解くことができる

方法１）EXCEL のソルバーアドイン(solver)を使用

方法２）フリーソフト R を使用

方法３）最適化パッケージ Numerical Optimizer を使用

(2) $W^{-l_1} + W^{-l_2} + \cdots + W^{-l_n} = 1$　について正根 W_0 を求める方法は、最後に提示する。

以下の５例を考えて各種方法で計算し、計算結果の一致を検証してみよう。

［例１］電話線障害発生率

９変数　$l_i = i$

［例２］市外電話の呼量

３変数　$l_i = i$

［例３］日本信販のクーポン

２変数　$l_i = 1,\ l_2 = 5$

［例４］ピアノの販売比率

２変数　$l_1 = 7,\ l_2 = 9$

［例５］家庭用電燈ランク構成比率

４変数　$l_1 = 1,\ l_2 = 2,\ l_3 = 4,\ l_4 = 6$

註1 エントロピー・モデル　國澤清典　日科技連

１． $H/\bar{l}$　最大化問題として直接解く

モデル

変数　p_i　，　$0 \le p_i \le 1$

$$\sum_i p_i = 1$$

評価値　l_i

目的関数　$H/\bar{l}$　最大化

ここで,

$$H = -\sum_i p_i \log p_i$$

$$\bar{l} = \sum_i l_i p_i$$

１．１　EXCEL のソルバーアドイン(solver)を使用

［例５］家庭用電燈ランク構成比率

４変数　$l_1 = 1,\ l_2 = 2,\ l_3 = 4,\ l_4 = 6$

で説明する。

固定部分入力（表１―１）

(1) 評価値(li)　$B$1:$E$1

(2) 変数(pi)　初期値設定(=1/n) ここが解になる　$B$2:$E$2

(3) 固定値　1　$C$3

表１－１　固定部分

	A	B	C	D	E	
1	li	1	2	4	6	
2	pi	0.25	0.25	0.25	0.25	
3	sum(pi)		1			
4	li*pi					
5	pi*Log(pi)					
6	lbar					
7	H					
8	H/lbar					

数式入力（表１－２）

(4) sum(pi)行　$B$3　=SUM(B2:E2)

(5) li*pi 行　$B$4　=B1*B2，　C,D,E 同様

(6)pi*log(pi)行　$B$5　=IF(B2>0,B2*LOG(B2),0)，　C,D,E 同様

(7)lbar 行　$B$6　=SUM(B4:E4)

(8)H 行　$B$7　=-SUM(B5:E5)

(9)H/lbar　$B$8　=B7/B6　目的関数値

表１－２　数式入力

	A	B	C	D	E
1	li	1	2	4	6
2	pi	0.25	0.25	0.25	0.25
3	sum(pi)	1	1		
4	li*pi	0.25	0.5	1	1.5
5	pi*Log(pi)	-0.15051	-0.15051	-0.15051	-0.15051
6	lbar	3.25			
7	H	0.60206			
8	H/lbar	0.185249			

ソルバーのパラメーターへの指定

(1) 目的セルの設定:(T)　$B$8

(2) 目標値　最大値(M) 選択

(3) 変数セルの変更:(B)　$B$2:$E$2

(4) 制約条件の対象:(U)　$B$3 = $C$3

(5) 制約のない変数を非負数にする(K) をチェックする

(6) 解決方法の選択:(E)　GRG 非線形　を選択

(7) 解決:(K)ボタンをクリックして計算する。

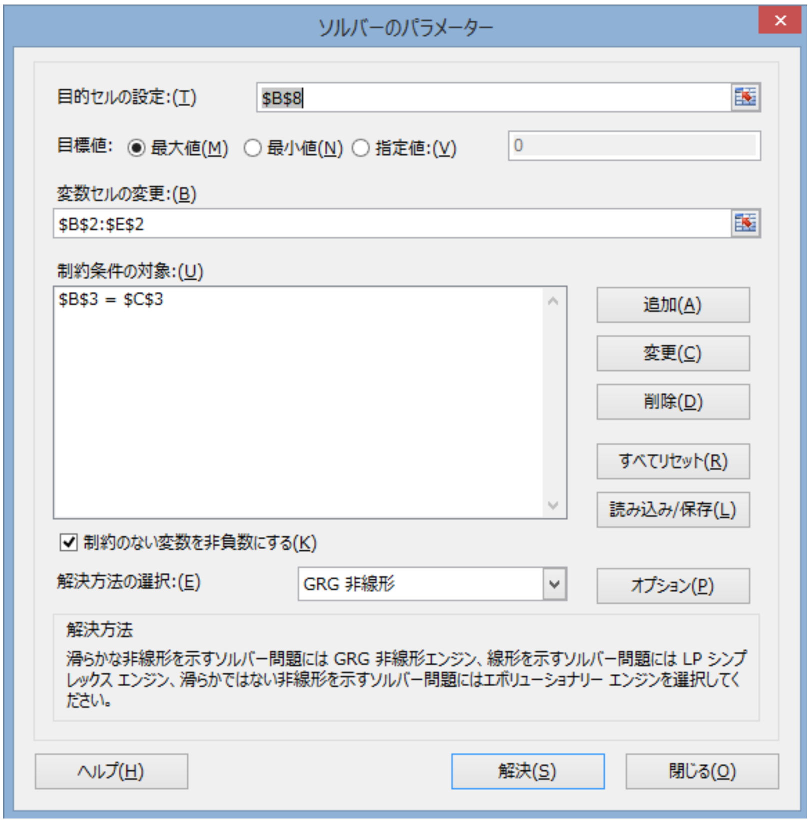

表１－３　結果　黄色部分が解、緑色部分が目的関数値

	A	B	C	D	E
1	li	1	2	4	6
2	pi	0.559115	0.31261	0.097725	0.03055
3	sum(pi)	1	1		
4	li*pi	0.559115	0.62522	0.3909	0.183299
5	pi*Log(pi)	−0.14118	−0.15787	−0.0987	−0.04628
6	lbar	1.758534			
7	H	0.444027			
8	H/lbar	0.252499			

変数 p_i の値が，本と一致していることがわかる。

以下[例１]から[例４]までも一致。

表１－４　［例１］電話線障害発生率

	A	B	C	D	E	F	G	H	I	J
1	li	1	2	3	4	5	6	7	8	9
2	pi	0.50049	0.25049	0.12537	0.06275	0.03141	0.01572	0.00787	0.00394	0.00197
3	sum(pi)	1	1							
4	li*pi	0.50049	0.50098	0.37611	0.25099	0.15703	0.09431	0.05507	0.0315	0.01773
5	pi*Log(pi)	−0.1504	−0.1506	−0.1131	−0.0754	−0.0472	−0.0283	−0.0166	−0.0095	−0.0053
6	lbar	1.98421								
7	H	0.59646								
8	H/lbar	0.3006								

表１－５　［例２］市外電話の呼量

	A	B	C	D
1	li	1	2	3
2	pi	0.54367839	0.29561089	0.16071072
3	sum(pi)	1	1	
4	li*pi	0.54367839	0.59122178	0.48213217
5	pi*Log(pi)	−0.1438888	−0.1564608	−0.1275971
6	lbar	1.61703234		
7	H	0.42794671		
8	H/lbar	0.26464944		

表１－６　［例３］日本信販のクーポン

	A	B	C
1	li	1	5
2	pi	0.75487799	0.24512284
3	sum(pi)	1.00000083	1
4	li*pi	0.75487799	1.2256142
5	pi*Log(pi)	−0.0921881	−0.149676
6	lbar	1.9804922	
7	H	0.24186412	
8	H/lbar	0.12212324	

表１－７ ［例４］ピアノの販売比率

	A	B	C
1	li	7	9
2	pi	0.54344871	0.45655129
3	sum(pi)	1	1
4	li*pi	3.80414094	4.10896165
5	pi*Log(pi)	−0.1439277	−0.1554605
6	lbar	7.91310259	
7	H	0.29938821	
8	H/lbar	0.03783449	

この方法で一番重要な点は，pi*log(pi)の計算で pi がゼロの時の扱いで、
無条件では，計算途中での pi がゼロで計算不可になってしまう。
そこで、IF(B2>0,B2*LOG(B2),0) のように pi>0 の条件を追加する。この条件がないと対象数が４以上で初期値をうまく設定しないと、計算不可になる場合が多く発生する。

１．２ フリーソフト R[註2]を使用

Rsolnp[註3]というライブラリを使用すると R で非線形制約条件付非線形最適化を行うことができるということなので，試してみた。

ライブラリー内の solnp 関数を使用。この関数の詳細は、註３の Rsolnp.pdf を参照のこと。
この関数を使うには、変数初期値、目的関数、等号制約関数、等号制約の値、不等号制約関数、不等号制約の下限値、上限値、変数の下限値、上限値等を指定する。
今回のモデルでは、等号制約が１つで、以下の関数を作成して solnp 関数に引き渡す。

```
# 目的関数 solnp は最小化なので、関数値の符号を逆にして使用する。
func.h <- function (pi, li)
{
        h <- -sum(pi*log(pi))
        lbar <- sum(pi*li)
        return(-h/lbar)
}
# 等号制約関数
func.p <-function (pi, li)
{
        return(sum(pi))
}
```

註2 R は、統計計算とグラフィックのためのフリーソフト http://www.r-project.org/
註3 http://cran.r-project.org/web/packages/Rsolnp/Rsolnp.pdf

solnp 関数は、多種な情報を戻すが、ここでは解($pars)のみ表示させる。
各例の評価値 l_i は、solnp 関数にベクトル形式で引数に追加する。
初期値は、すべて同じ値(1/n)にしておく。

［例１］電話線障害発生率

９変数 $l_i = i$

```
solnp(pars=rep(1/9,9), fun=func.h, eqfun=func.p, eqB = c(1), li=c(1:9))$pars
[1] 0.500494103 0.250492838 0.125369827 0.062746799 0.031404355 0.015717734
[7] 0.007866629 0.003937201 0.001970515
```

［例２］市外電話の呼量

３変数 $l_i = i$

```
solnp(pars=rep(1/3,3), fun=func.h, eqfun=func.p, eqB = c(1), li=c(1,2,3))$pars
[1] 0.5436890 0.2955978 0.1607132
```

［例３］日本信販のクーポン

２変数 $l_i = 1,\ l_2 = 5$

```
solnp(pars=c(0.5,0.5), fun=func.h, eqfun=func.p, eqB = c(1), li=c(1,5))$pars
[1] 0.7548777 0.2451223
```

［例４］ピアノの販売比率

２変数 $l_1 = 7,\ l_2 = 9$

```
solnp(pars=c(0.5,0.5), fun=func.h, eqfun=func.p, eqB = c(1), li=c(7,9))$pars
[1] 0.5434487 0.4565513
```

［例５］家庭用電燈ランク構成比率

４変数 $l_1 = 1,\ l_2 = 2,\ l_3 = 4,\ l_4 = 6$

```
solnp(pars=rep(1/4,4), fun=func.h, eqfun=func.p, eqB = c(1), li=c(1,2,4,6))$pars
[1] 0.55911531 0.31260991 0.09772500 0.03054978
```

以上、いずれも國澤先生の本の値と一致している。

１．３　最適化パッケージ Numerical Optimizer[註4] を使用

モデル　one_factor.smp

```
Set P;
Element i(set=P);
Variable p(name="p", index=i);
```

[註4] 株式会社　NTT データ数理システムで販売している最適化パッケージ

```
Parameter l(name="l", index=i);
Expression H(name="H");
Expression lu(name="lu");

0 <= p[i] <= 1;
sum(p[i], i) == 1;
p[i] = 1.0/P.card();

H = -sum(p[i]*log10(p[i]), i);
lu = sum(p[i]*l[i], i);

Objective H1(type=maximize);
H1 = H/lu;

solve();

p[i].val.print();
H.val.print();
lu.val.print();
H1.val.print();
```

［例１］電話線障害発生率

９変数　$l_i = i$

データ　　rei1.dat

```
l = [1] 1 [2] 2 [3] 3 [4] 4 [5] 5 [6] 6 [7] 7 [8] 8 [9] 9;
>one_factor rei1.dat
…
ELAPSED_TIME(sec.)                                0.09
SOLUTION_FILE                           one_factor.sol
p[1]=0.500493
p[2]=0.250493
p[3]=0.12537
p[4]=0.0627469
p[5]=0.0314044
p[6]=0.0157177
p[7]=0.00786663
```

p[8]=0.00393721

p[9]=0.00197056

H=0.596456

lu=1.98421

H1=0.300602

［例２］市外電話の呼量

３変数　$l_i = i$

データ　　rei2.dat

l = [1] 1 [2] 2 [3] 3 ;

>one_factor rei2.dat

…

```
ELAPSED_TIME(sec.)                                          0.05
SOLUTION_FILE                                     one_factor.sol
p[1]=0.543689
p[2]=0.295598
p[3]=0.160713
H=0.427945
lu=1.61702
H1=0.264649
```

［例３］日本信販のクーポン

２変数　$l_i = 1,\ l_2 = 5$

データ　　rei3.dat

l = [1] 1 [2] 5;

>one_factor rei3.dat

…

ELAPSED_TIME(sec.)　　0.03

SOLUTION_FILE　　one_factor.sol

p[1]=0.754878

p[2]=0.245122

H=0.241864

lu=1.98049

H1=0.122123

［例４］ピアノの販売比率

２変数　$l_1 = 7,\ l_2 = 9$

データ　　rei4.dat

l = [1] 7 [2] 9;

```
>one_factor rei4.dat
…
ELAPSED_TIME(sec.)                                          0.04
SOLUTION_FILE                                      one_factor.sol
p[1]=0.543449
p[2]=0.456551
H=0.299388
lu=7.9131
H1=0.0378345
```

［例５］家庭用電燈ランク構成比率

４変数　$l_1 = 1,\ l_2 = 2,\ l_3 = 4,\ l_4 = 6$

データ　rei5.dat

```
l = [1] 1 [2] 2 [3] 4 [4] 6;
```

```
>one_factor rei5.dat
…
ELAPSED_TIME(sec.)                                          0.05
SOLUTION_FILE                                      one_factor.sol
p[1]=0.559115
p[2]=0.31261
p[3]=0.097725
p[4]=0.0305498
H=0.444027
lu=1.75853
H1=0.252499
```

２．$W^{-l_1} + W^{-l_2} + \cdots + W^{-l_n} = 1$　について正根W_0を求める方法

ここでは、統計解析パッケージ[註5]の uniroot 関数を使用して解を求める方法を説明する。uniroot 関数は，１変数の連続関数に対して、関数値がゼロとなるものを探す**ニュートン法を使用している。この方法に関しては，後述する。**

以下の root.W 関数を作成して使用する。

```
  root.W <-
function(w, l)
{
        f <- sum(w^( - l)) - 1
        f
}
```

以下 ［例１］～［例５］に関しての計算結果を示す。
すべて本の結果と一致しているのを確認されたい。

註5 S-PLUS （NTT データ数理システムが販売している）または R(フリーソフト)を使用

［例１］電話線障害発生率

９変数　$l_i = i$

```
uniroot(root.W, lower=1, upper=3, l=c(1:9))$root
[1] 1.998022
(uniroot(root.W, lower=1, upper=3, l=c(1:9))$root)^(-c(1:9))
[1] 0.500495027 0.250495273 0.125371638 0.062747882 0.031405003
[6] 0.015718048 0.007866805 0.003937297 0.001970597
```

［例２］市外電話の呼量

３変数　$l_i = i$

```
uniroot(root.W, lower=1, upper=3, l=c(1:3))$root
[1] 1.839286
(uniroot(root.W, lower=1, upper=3, l=c(1:3))$root)^(-c(1:3))
[1] 0.5436893 0.2955980 0.1607135
```

［例３］日本信販のクーポン

２変数　$l_i = 1,\ l_2 = 5$

```
uniroot(root.W, lower=1, upper=3, l=c(1,5))$root
[1] 1.324718
(uniroot(root.W, lower=1, upper=3, l=c(1,5))$root)^(-c(1,5))
[1] 0.7548779 0.2451227
```

［例４］ピアノの販売比率

２変数　$l_1 = 7,\ l_2 = 9$,

```
uniroot(root.W, lower=1, upper=3, l=c(7,9))$root
[1] 1.091024
(uniroot(root.W, lower=1, upper=3, l=c(7,9))$root)^(-c(7,9))
[1] 0.5434488 0.4565514
```

［例５］家庭用電燈ランク構成比率

４変数　$l_1 = 1,\ l_2 = 2,\ l_3 = 4,\ l_4 = 6$

```
uniroot(root.W, lower=1, upper=3, l=c(1,2,4,6))$root
[1] 1.78854
(uniroot(root.W, lower=1, upper=3, l=c(1,2,4,6))$root)^(-c(1,2,4,6))
[1] 0.55911531 0.31260993 0.09772497 0.03054980
```

ニュートン法

x の関数 f(x) について，f(x) = 0 を満たす解を求める方法をニュートン法という.
x のある範囲[a,b]において関数が x 軸と交わる点の x 座標を求める。
f(a) > 0, f(b) < 0　または　f(a) < 0, f(b) > 0 になるように範囲[a, b] を設定すれば解が計算出来る。(uniroot 関数の引数 lower, upper で a, b を指定)

$$f(x) = W^{-l_1} + W^{-l_2} + \cdots + W^{-l_n} - 1$$

$f(a)>0,\ f(b)<0$ となるような，a, b を決める。

$f(1)=1\char`^(-l_1)+1\char`^(-l_2)+\ldots+1-(-l_n)-1=1+1+\ldots+1-1=n-1$
$f(1)>0$ であるから，a= 1 が適当である。

b は，２～３で $f(b)<0$ になるが，
l の値が大きいと，W^{-l} の値は，$W=2$ の場合で，
l が５０を超えると，ほとんどゼロになり，正しい解は，求められない。
> 2^(-50)
[1] 8.881784e-016

そこで，l の値が大きい時には、値を調整（スケール化）する必要がある。
l_i での解を W_0，$L_i=l_i\cdot\alpha$ での解を W_1 とすると，
$W_o^{-l_i}=W_1^{-L_i}=W_1^{-l\cdot\alpha}=(W_1^{\alpha})^{-l_i}$ ==> $W_0=W_1^{\alpha}$
つまり　l_i の値をα倍して，W を計算し，α乗倍すれば，元の解が求まる。

$\max(l_i)\cdot\alpha=10$ となる，αを決めて　b=3 を採用すれば，精度の良い解が求められる。