

# 最尤法について

????,050610, 1111,18,121003,20210513, 27, 0610 2021. 6. 10()

## 1 最尤解

$Y_i, i = 1, 2, \dots, n$  を独立な観測値とする。 $f(y_i|\theta)$  をデータの確率関数（密度関数）とする。データの同時分布の確率関数は

$$f(y_1, y_2, \dots, y_n|\theta) = \prod_{i=1}^n f(y_i|\theta) \quad (1)$$

である。これは、母数の値が  $\theta$  の時に、 $y_1, y_2, \dots, y_n$  というデータを観測する確率（密度）であり、その値が高いほど、それらのデータ系列（パターン）が出現する確率が高い。

この同時確率関数を母数  $\theta$  の関数と見たものが尤度関数である。

$$L(\theta) = f(y_1, y_2, \dots, y_n|\theta) \quad (2)$$

これは、データ系列を与えられた時に、母数の値がどの程度尤もらしいかを表す値と解釈することが出来る。ただし、 $\theta$  の確率を与えるものではない。（ $\theta$  は確率変数ではないので確率は付与できない。）

なお、パラメタを含まない部分は、尤度関数の定義から落としても良い。その意味で、

$$L(\theta) \propto f(y_1, y_2, \dots, y_n|\theta) \quad (3)$$

と書く場合がある。

尤度関数の対数を対数尤度関数という。

$$\ln L(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(y_i|\theta) \quad (4)$$

多くの場合、対数尤度関数の方が尤度関数よりも簡単な形をしている。

尤度関数もしくは対数尤度関数の最大値を与える母数  $\theta$  の値、 $\hat{\theta}$ 、をその母数の推定値とする方法を、最尤解と呼ぶ。対数関数は単調増加関数であるため、対数尤度関数の最大値を与える母数はもとの尤度関数の最大値を与える。

母数が連続量の場合、最尤解を求めるためには対数尤度関数の母数  $\theta$  に関する微分を 0 とおいた方程式

$$d \ln L(\theta) = \frac{d \ln L(\theta)}{d \theta} = 0 \quad (5)$$

を解けばよい。なお、得られた解で、対数尤度の 2 次微分が負になっていることが必要である。

### 1.1 正規分布の場合（ $\sigma$ は既知でそれを 1 と仮定）

$$f(y_1, y_2, \dots, y_n|\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i - \theta)^2\right) \quad (6)$$

$$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i - \theta)^2\right) \quad (7)$$

$$\propto \prod_{i=1}^n \exp\left(-\frac{1}{2}(y_i - \theta)^2\right) \quad (8)$$

$$\ln L(\theta) = \sum_{i=1}^n -\frac{1}{2}(y_i - \theta)^2 + \text{const} \quad (9)$$

$$d \ln L(\theta) = \sum_{i=1}^n (y_i - \theta) = 0 \quad (10)$$

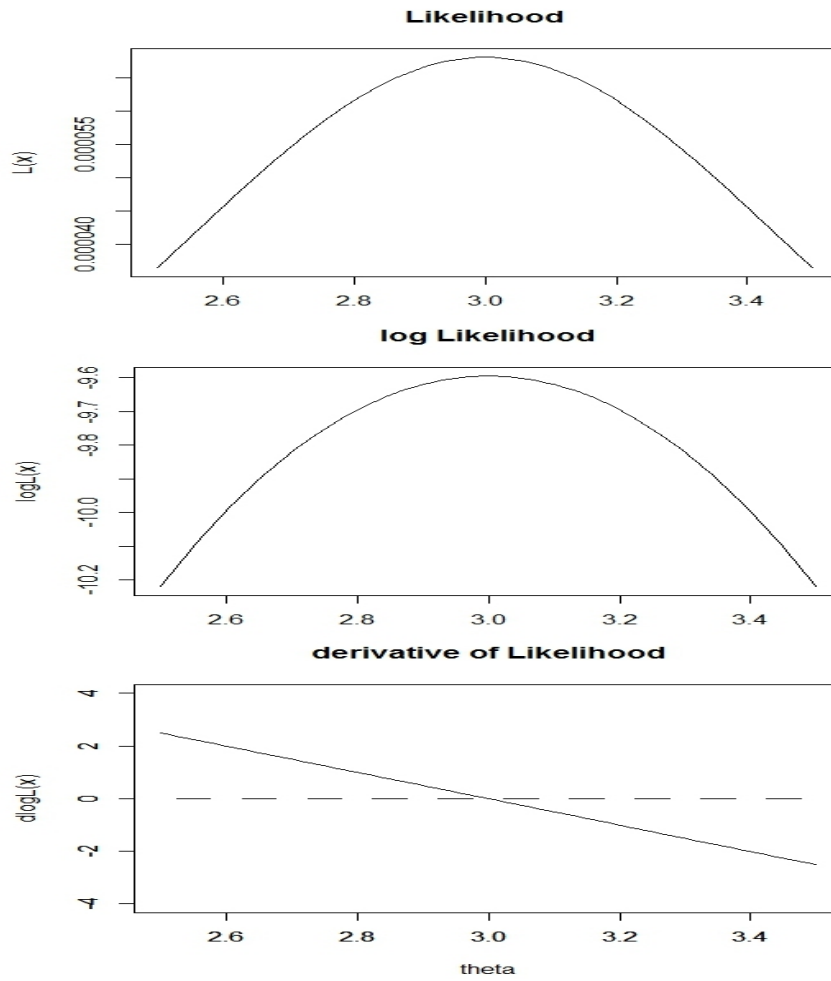
したがって、

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n y_i \quad (11)$$

### 1.1.1 数値例 (正規分布)

$y = \{1, 2, 3, 4, 5\}$  の場合

THETA	L	LL	DLL
2.5	0.000019	-10.8236	2.5000
2.6	0.000025	-10.5986	2.0000
2.7	0.000029	-10.4236	1.5000
2.8	0.000033	-10.2986	1.0000
2.9	0.000036	-10.2236	0.5000
3	0.000037	-10.1986	0.0000 *
3.1	0.000036	-10.2236	0.5000
3.2	0.000033	-10.2986	1.0000
3.3	0.000029	-10.4236	1.5000
3.4	0.000025	-10.5986	2.0000
3.5	0.000019	-10.8236	2.5000



## 1.2 ベルヌーイ分布の場合

$$f(y_1, y_2, \dots, y_n | \theta) = \prod_{i=1}^n \theta^{y_i} (1 - \theta)^{1-y_i} \quad (12)$$

$$L(\theta) = \prod_{i=1}^n \theta^{y_i} (1 - \theta)^{1-y_i} \quad (13)$$

$$\ln L(\theta) = \sum_{i=1}^n (y_i \log \theta + (1 - y_i) \log(1 - \theta)) \quad (14)$$

$$d \ln L(\theta) = \sum_{i=1}^n \left( \frac{y_i}{\theta} - \frac{(1 - y_i)}{(1 - \theta)} \right) = 0 \quad (15)$$

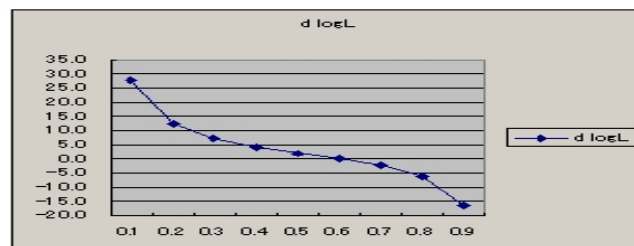
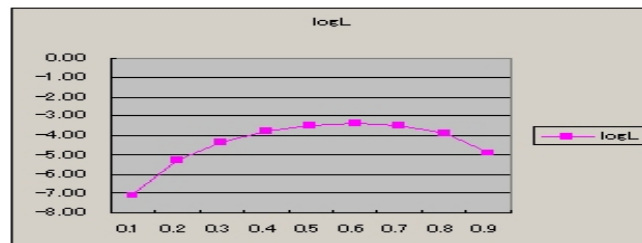
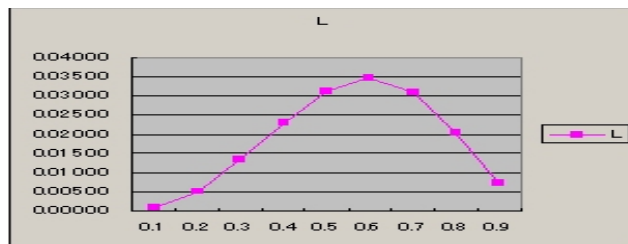
したがって、

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n y_i \quad (16)$$

### 1.2.1 数値例（2項分布）

$y = \{0, 1, 1, 0, 1\}$  の場合

THETA	L	LL	DLL
0.1	0.0008	-7.1185	27.7778
0.2	0.0051	-5.2746	12.5000
0.3	0.0132	-4.3253	7.1429
0.4	0.0230	-3.7705	4.1667
0.5	0.0313	-3.4657	2.0000
0.6	0.0346	-3.3651	-0.0000 *
0.7	0.0309	-3.4780	-2.3810
0.8	0.0205	-3.8883	-6.2500
0.9	0.0073	-4.9213	-16.6667



### 1.3 正規分布の場合 ( $\mu, \sigma$ とともに未知)

母数  $\theta$  を

$$\tau = g(\theta) \quad (17)$$

と変換し新たな母数  $\tau$  を作ったとすると、対数尤度の最大値を与える  $\theta$  の値は対数尤度の最大値を与える  $\tau$  を

$$\theta = g^{-1}(\tau) \quad (18)$$

と変換したものとなる。

以下に

$$\tau = 1/\sigma^2 \quad (19)$$

を用いる。

$$f(y_1, y_2, \dots, y_n | \mu, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}(y_i - \mu)^2/\sigma^2\right) \quad (20)$$

$$L(\mu, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}(y_i - \mu)^2/\sigma^2\right) \quad (21)$$

$$\propto \left(\frac{1}{\sigma}\right)^n \prod_{i=1}^n \exp\left(-\frac{1}{2}(y_i - \mu)^2/\sigma^2\right) \quad (22)$$

$$\ln L(\mu, \sigma) = -n \ln \sigma + \sum_{i=1}^n -\frac{1}{2}(y_i - \mu)^2/\sigma^2 + const \quad (23)$$

$$= -\frac{1}{2} \left( n \ln \sigma^2 + \sum_{i=1}^n (y_i - \mu)^2/\sigma^2 \right) + const \quad (24)$$

$$= -\frac{1}{2} \left( -n \ln \tau + \sum_{i=1}^n (y_i - \mu)^2 \tau \right) + const \quad (25)$$

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^n (y_i - \mu)/\sigma^2 = \sum_{i=1}^n (y_i - \mu) \tau \quad (26)$$

$$\frac{\partial \ln L}{\partial \sigma} = ??? \quad (27)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = ??? \quad (28)$$

$$\frac{\partial \ln L}{\partial \tau} = -\frac{n}{\tau} + \sum_{i=1}^n (y_i - \mu)^2 \quad (29)$$

したがって、

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i \quad (30)$$

$$\hat{\tau} = \frac{n}{\sum_{i=1}^n (y_i - \hat{\mu})^2} \quad (31)$$

$$\widehat{\sigma^2} = \frac{1}{\hat{\tau}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu})^2 \quad (32)$$

$$\hat{\sigma} = \sqrt{\widehat{\sigma^2}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu})^2} \quad (33)$$

## 2 回帰分析

### 2.1 正規分布の場合：回帰分析

データ

$$(x_i, y_i), i = 1, 2, \dots, n \quad \text{ただし } y_i \text{ は連続量} \quad (34)$$

モデル

$$y_i = \alpha x_i + \beta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \quad (35)$$

または

$$\mathcal{E}(y_i | x_i) = \hat{y}_i = \alpha x_i + \beta \quad (36)$$

モデルのパラメタは  $\xi = \{\alpha, \beta, (\sigma^2)\}$

尤度関数

$$L(\xi) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(y_i - (\alpha x_i + \beta))^2 / \sigma^2\right) \quad (37)$$

対数尤度関数

$$\ln L(\xi) = -\frac{1}{2} \left( n \ln(\sigma^2) + \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - (\alpha x_i + \beta))^2 \right) + \text{const} \quad (38)$$

これを最大にするように  $\xi$  を求める。すなわち、連立方程式

$$\frac{\partial \ln L}{\partial \alpha} = 0, \quad \frac{\partial \ln L}{\partial \beta} = 0, \quad \frac{\partial \ln L}{\partial \sigma} = 0 \quad (39)$$

を解く。答えは

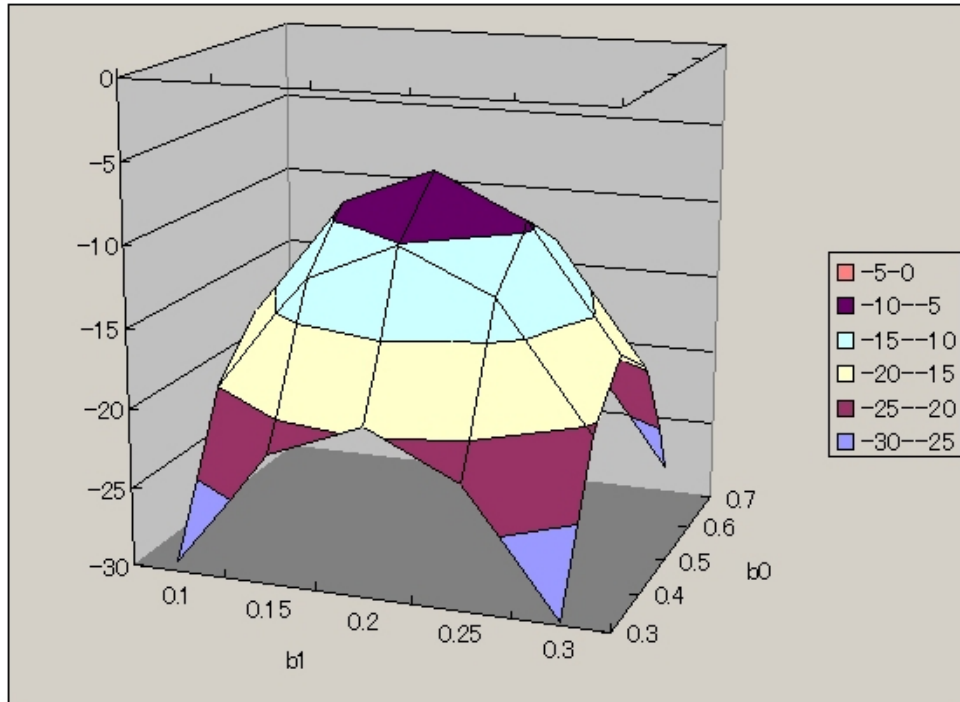
$$\hat{\alpha} = \frac{\text{cov}(x, y)}{\text{var}(x)}, \quad \hat{\beta} = \bar{y} - \hat{\alpha} \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - (\hat{\alpha} x_i + \hat{\beta}))^2 \quad (40)$$

Data1				最尤解の探索		
Obs	X	y	yhat	alpha	beta	LL
				0.3	0.1	-29.316
1	-3.0	0.000	-0.200	0.3	0.15	-21.841
2	-2.6	0.000	-0.080	0.3	0.2	-19.438
3	-2.3	0.000	0.010	0.3	0.25	-22.105
4	-1.9	0.000	0.130	0.3	0.3	-29.843
5	-1.5	0.200	0.250	0.4	0.1	-19.984
6	-1.1	0.166	0.370	0.4	0.15	-12.509
7	-0.8	0.285	0.460	0.4	0.2	-10.106
8	-0.4	0.125	0.580	0.4	0.25	-12.773
9	0.0	0.777	0.700	0.4	0.3	-20.511
10	0.4	0.875	0.820	0.5	0.1	-16.614
11	0.8	0.714	0.940	0.5	0.15	-9.139
12	1.1	1.000	1.030	0.5	0.2	-6.735 *
13	1.5	0.800	1.150	0.5	0.25	-9.402
14	1.9	1.000	1.270	0.5	0.3	-17.141
15	2.3	0.666	1.390	0.6	0.1	-19.205
16	2.6	1.000	1.480	0.6	0.15	-11.730
17	3.0	1.000	1.600	0.6	0.2	-9.326
				0.6	0.25	-11.994
				0.6	0.3	-19.732
				0.7	0.1	-27.757
				0.7	0.15	-20.283
				0.7	0.2	-17.879
				0.7	0.25	-20.546
				0.7	0.3	-28.284

**最尤解**

$\beta$  0.5065  
 $\alpha$  0.1987  
 $\sigma$  0.1689

ただし、表中 yhat は上記の最尤解で計算したもの。



## 2.2 2項分布の場合：ロジスティック回帰分析

データ

$$(x_i, y_i), i = 1, 2, \dots, n \quad \text{ただし } y_i \text{ は } 0 \text{ か } 1 \quad (41)$$

これを  $x_i$  の異なる値ごとに度数分布の形にしたもの

$$(x_i, f_i, r_i), i = 1, 2, \dots, N \quad \text{ただし } f_i \text{ は試行数 } r_i \text{ は成功数} \quad (42)$$

モデル

確率は  $(0, 1)$  の区間に入らなければならないので、

$$\Pr(Y_i = 1|x_i) = \frac{1}{1 + \exp(-(\alpha x_i + \beta))} = P(x_i|\alpha, \beta) = \pi_i \quad (43)$$

とする。これは

$$\log \frac{\Pr(Y_i = 1|x_i)}{1 - \Pr(Y_i = 1|x_i)} = \alpha x_i + \beta \quad (44)$$

と同値である。なお、モデルのパラメタは  $\xi = \{\alpha, \beta\}$  であり、

$$\mathcal{E}(Y_i|x_i) = P(x_i|\alpha, \beta) = \pi_i \quad (45)$$

$$\text{var}(Y_i|x_i) = \pi_i(1 - \pi_i) \quad (46)$$

となる。また、 $(f_i, r_i)$  を試行数と度数として、

$$\mathcal{E}(R_i|x_i) = f_i \times P(x_i|\alpha, \beta) = f_i \pi_i \quad (47)$$

$$\text{var}(R_i|x_i) = f_i \pi_i(1 - \pi_i) \quad (48)$$



である。

尤度関数

$$L(\boldsymbol{\xi}) = \prod_{i=1}^n P(x_i|\alpha, \beta)^{y_i} (1 - P(x_i|\alpha, \beta))^{1-y_i} \quad (49)$$

$$L(\boldsymbol{\xi}) = \prod_{i=1}^n P(x_i|\alpha, \beta)^{r_i} (1 - P(x_i|\alpha, \beta))^{f_i-r_i} \quad (50)$$

対数尤度関数

$$\ln L(\boldsymbol{\xi}) = \sum_{i=1}^n y_i \log P(x_i|\alpha, \beta) + (1 - y_i) \log(1 - P(x_i|\alpha, \beta)) \quad (51)$$

$$\ln L(\boldsymbol{\xi}) = \sum_{i=1}^n r_i \log P(x_i|\alpha, \beta) + (f_i - r_i) \log(1 - P(x_i|\alpha, \beta)) \quad (52)$$

これを最大にするように  $x_i$  を求める。すなわち、連立方程式

$$\frac{\partial \ln L}{\partial \alpha} = 0, \quad \frac{\partial \ln L}{\partial \beta} = 0, \quad (53)$$

を解く。答えは解析的には解けない。

IRT のパラメタとの関係

指数関数の中が、IRT の場合は  $-1.7a(x - b)$ , LR:  $-(\alpha x + \beta)$  であるから、 $a = \alpha/1.7$ ,  $b = -\beta/\alpha$  となる。

例

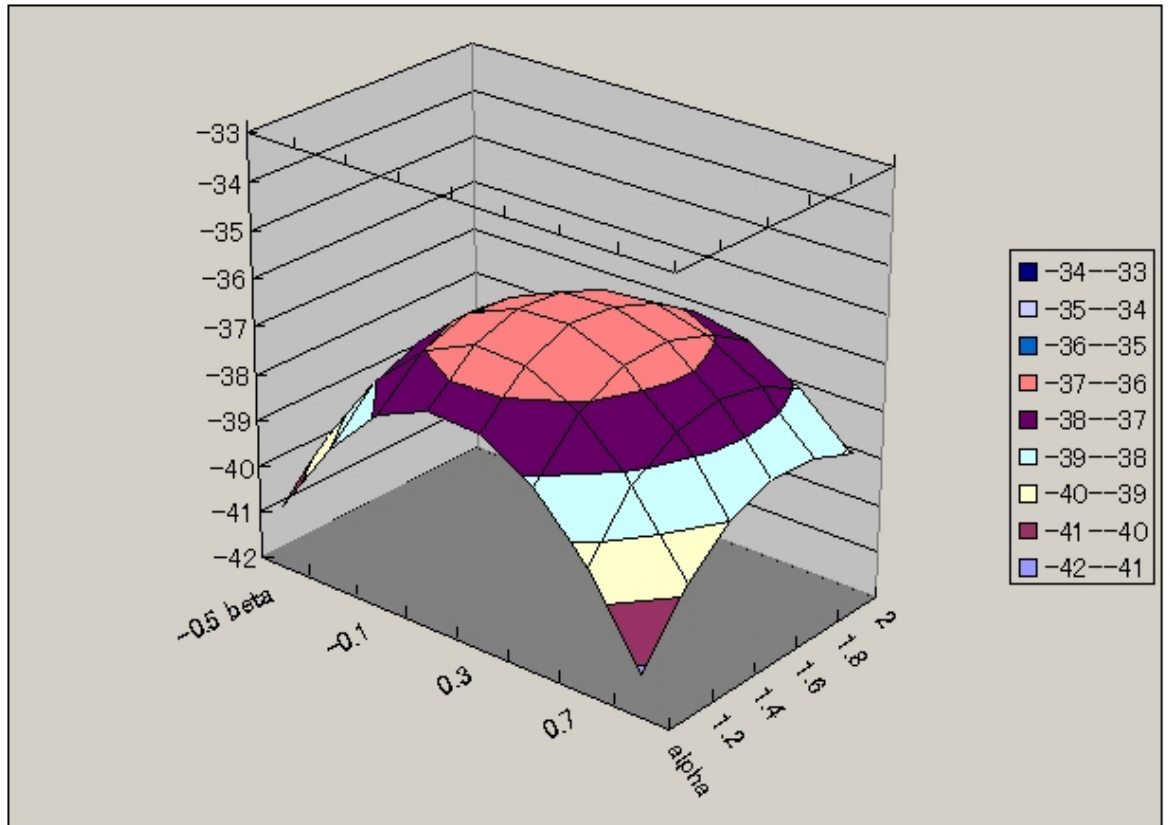
Data1						最尤解の探索		
Obs	X	F	R	P	phat	ALPHA	BETA	LL
						1.0	-0.5	-40.6408
1	-3.0	1	0	0.000	0.015	1.0	-0.3	-38.9813
2	-2.6	2	0	0.000	0.026	1.0	-0.1	-37.8991
3	-2.3	3	0	0.000	0.040	1.0	0.1	-37.3991
4	-1.9	4	0	0.000	0.071	1.0	0.3	-37.4813
5	-1.5	5	1	0.200	0.120	1.0	0.5	-38.1408
6	-1.1	6	1	0.166	0.198	1.0	0.7	-39.3681
7	-0.8	7	2	0.285	0.277	1.0	0.9	-41.1493
8	-0.4	8	1	0.125	0.407	1.2	-0.5	-39.6113
9	0.0	9	7	0.777	0.552	1.2	-0.3	-38.0549
10	0.4	8	7	0.875	0.689	1.2	-0.1	-37.0249
11	0.8	7	5	0.714	0.799	1.2	0.1	-36.5249
12	1.1	6	6	1.000	0.860	1.2	0.3	-36.5549
13	1.5	5	4	0.800	0.917	1.2	0.5	-37.1113
14	1.9	4	4	1.000	0.952	1.2	0.7	-38.1873
15	2.3	3	2	0.666	0.972	1.2	0.9	-39.7725
16	2.6	2	2	1.000	0.982	1.4	-0.5	-39.1354
17	3.0	1	1	1.000	0.990	1.4	-0.3	-37.6707
						1.4	-0.1	-36.6870
						1.4	0.1	-36.1870
						1.4	0.3	-36.1707 *
						1.4	0.5	-36.6354
						1.4	0.7	-37.5759
						1.4	0.9	-38.9847
						1.6	-0.5	-39.0962
						1.6	-0.3	-37.7122
						1.6	-0.1	-36.7692
						1.6	0.1	-36.2692
						1.6	0.3	-36.2122
						1.6	0.5	-36.5962
						1.6	0.7	-37.4173
						1.6	0.9	-38.6698
						1.8	-0.5	-39.4032
						1.8	-0.3	-38.0898
						1.8	-0.1	-37.1823
						1.8	0.1	-36.6823
						1.8	0.3	-36.5898
						1.8	0.5	-36.9032
						1.8	0.7	-37.6195
						1.8	0.9	-38.7345
						2.0	-0.5	-39.9865
						2.0	-0.3	-38.7348
						2.0	-0.1	-37.8585
						2.0	0.1	-37.3585

ただし、 phat は最尤推定値で計算したもの。

2.0	0.3	-37.2348
2.0	0.5	-37.4865
2.0	0.7	-38.1110
2.0	0.9	-39.1051

**最优解**

$\alpha$  1.4642 (0.861)  
 $\beta$  0.2127 (-0.14)



### 3 数値解

一般に、ある関数  $r(\xi)$  の極値（極小値または極大値のことであるが、本節では極小値とする）を与えるパラメタ  $\xi$  の値を数値的に求めることを関数の最適化と呼ぶが、その方法は、何らかの方法で求めた  $\xi$  の初期値から、 $r(\xi)$  が減少する方向  $d$  へある分量  $\alpha$  だけ修正する、

$$\xi_{(\ell+1)} = \xi_{(\ell)} + \alpha d_{(\ell)} \quad (54)$$

という形の繰り返し計算でパラメタの値を更新していくという形をとる。なお、この関数  $r(\xi)$  は目的関数 objective function と呼ばれる。

数値的最適化の方法には様々なものがあるが、それらの主な違いは、更新の方向  $d_{(\ell)}$  をどの様にして定めるかによる。一般的には、以下に示すように、 $g(\xi_{(\ell)})$  を  $r(\xi)$  のパラメタの各要素に関する一次偏微分係数ベクトル (gradient) を  $\xi = \xi_{(\ell)}$  で評価したものとして、

$$g(\xi_{(\ell)})' d_{(\ell)} < 0 \quad (55)$$

であれば

$$r(\xi_{(\ell+1)}) \leq r(\xi_{(\ell)}) \quad (56)$$

を満たすステップサイズ  $\alpha$  が存在する。

#### 3.1 最急降下法 method of steepest descent

最も単純なものは最急降下法と呼ばれる方法であり、修正の方向を  $r(\xi)$  の gradient  $g$  (gradient) の逆方向

$$d_{(\ell)} = -g(\xi_{(\ell)})$$

に取るものである。ただし、

$$g(\xi_{(\ell)}) = \left. \frac{\partial r(\xi)}{\partial \xi} \right|_{\xi=\xi_{(\ell)}} \quad (57)$$

は  $r(\xi)$  の gradient を  $\xi = \xi_{(\ell)}$  で評価したものである。すなわち最急降下法の更新式は

$$\xi_{(\ell+1)} = \xi_{(\ell)} - \alpha g(\xi_{(\ell)}) \quad (58)$$

である。

この方法は単純であるが、関数の極値に到達するまでの繰り返しの数が多くかかる場合が多いことが欠点である。

##### 3.1.1 何故負の gradient 方向なのか

いま、 $r(\xi_{(\ell+1)})$  を  $\xi_{(\ell)}$  の周りで1次のテイラー展開をすると

$$r(\xi_{(\ell+1)}) \approx r(\xi_{(\ell)}) + g(\xi_{(\ell)})' (\xi_{(\ell+1)} - \xi_{(\ell)}) \quad (59)$$

であるが、これに方向を示す長さが 1 のベクトル  $\delta$  を用いて

$$\xi_{(\ell+1)} = \xi_{(\ell)} + \delta \quad (60)$$

を代入すると

$$r(\xi_{(\ell)} + \delta) \approx r(\xi_{(\ell)}) + g(\xi_{(\ell)})' \delta \quad (61)$$

を得る。したがって、 $r(\xi_{(\ell)} + \delta)$  を  $r(\xi_{(\ell)})$  より減少させるためには

$$g(\xi_{(\ell)})' \delta < 0 \quad (62)$$

である必要がある。(以下、 $g(\xi_{(\ell)})$  を  $g$  と記す。) また、 $r(\xi_{(\ell)} + \delta)$  を最小とするためには  $g' \delta$  を最小とするような長さが 1 のベクトル  $\delta$  を選べば良い。ここで、ベクトルの内積  $g' \delta$  に関して、以下の不等式が成り立つ。

$$-\|g\| \|\delta\| \leq g' \delta \leq \|g\| \|\delta\| \quad (63)$$

(ベクトルの内積とベクトル間の角度のコサインの関係、もしくは Cauchy-Schwarz の不等式を用いる。) そして、長さが 1 のベクトルは任意のベクトル  $m$  を用いて

$$\delta = \frac{m}{\|m\|} \quad (64)$$

と書けること、ならびに  $\|\delta\| = 1$  から、

$$-\|g\| \leq g' \frac{m}{\|m\|} \leq \|g\| \quad (65)$$

を得る。ここで  $\|g\|^2 = g'g$  であるから

$$m = -g \quad \text{したがって} \quad \delta = -\frac{g}{\|g\|} \quad (66)$$

の時に上式の左側の等号が成立する。すなわち、 $\xi_{(\ell)}$  から長さ 1 だけ移動する場合、Eq.(62) が示すように  $g$  との間の成す角が 90 度以内の任意のベクトル (ベクトル間のコサインが正) の符号を逆転したものを  $\delta$  として用いれば目的関数の値は減少する可能性があるが、特に、負の gradient の方向へ移動すれば  $\xi_{(\ell)}$  付近での目的関数の減少幅は最大となる。(局所的には最適であるが大局的に見ると必ずしもそうでは無い可能性がある。)

### 3.2 ニュートンラフソン法 The Newton-Raphson

一次微分のみならず、2次微分の情報をも用いて最適化を行う方法が、通常 Newton-Raphson 法と呼ばれる方法である。  $r(\xi)$  を  $\xi_{(\ell)}$  のまわりで2次の項まで Taylor 展開すると

$$r(\xi) \approx r(\xi_{(\ell)}) + g(\xi_{(\ell)})'(\xi - \xi_{(\ell)}) + \frac{1}{2}(\xi - \xi_{(\ell)})' H(\xi_{(\ell)})(\xi - \xi_{(\ell)})$$

となるが、この  $p \times p$  の2次微分行列 (  $r$  をパラメタの各要素で2回偏微分したものを要素として持つ行列)

$$H(\xi_{(\ell)}) = \left[ \frac{\partial^2 r(\xi)}{\partial \xi \partial \xi'} \Big|_{\xi = \xi_{(\ell)}} \right]$$

はヘシアン行列と呼ばれるものである。上記の表現 Eq.(67) は  $r(\xi)$  を二次関数で近似していることに他ならない。 $(\xi_{(\ell)})$  は定数であることに注意。

したがって、極値の条件は

$$\frac{\partial r(\xi)}{\partial \xi} \approx g(\xi_{(\ell)}) + H(\xi_{(\ell)})(\xi - \xi_{(\ell)}) = 0$$

となり、これより

$$\xi = \xi_{(\ell)} - H(\xi_{(\ell)})^{-1}g(\xi_{(\ell)})$$

となる。したがって、Eq.(54) において

$$d_{(\ell)} = -H(\xi_{(\ell)})^{-1}g(\xi_{(\ell)})$$

とおいたものが Newton-Raphson 法の更新式として得られる。Newton-Raphson 法は一般的に効率の良い方法であるが、ヘシアン行列の計算が複雑な場合には利用が困難であること、また、ヘシアン行列が必ずしも非負定値行列とはならないために、まれに Eq.(56) を満たすような  $\alpha$  を見いだすことが不可能な場合が存在する（負のグレイディエント方向から 90 度以上方向がそれる）等の欠点がある。このような欠点を補うために、1 次微分の情報のみからヘシアン行列の逆行列を近似的に計算する方法（いわゆる準ニュートン法）も考案されている。

### 3.2.1 一変数の場合

本節では  $\xi$  がスカラーの場合を考える。この場合、 $r(\xi)$  は一変数の関数となるが、それをスカラー量  $x$  を用いて  $f(x)$  と置く。この関数の  $x = x_0$  における 2 次までのテイラー展開は Eq.(67) より

$$f(x) \approx f^{(*)}(x) = f(x_0) + g(x_0)(x - x_0) + \frac{1}{2}h(x_0)(x - x_0)^2 \quad (67)$$

ただし

$$g(x) = \frac{df(x)}{dx} \quad \text{and} \quad h(x) = \frac{d^2f(x)}{dx^2} \quad (68)$$

という  $x$  の 2 次関数となる。 $(x_0)$  は定数であることに注意。

したがって、 $f^{(*)}(x)$  を微分して 0 と置いた方程式

$$\frac{df^{(*)}(x)}{dx} = g(x_0) + h(x_0)(x - x_0) = 0 \quad (69)$$

の解を

$$x_1 = x_0 - \frac{g(x_0)}{h(x_0)} \quad (70)$$

とすれば  $f^{(*)}(x)$  はそこで極値を取る。したがって、この近似を使えば、 $f(x)$  もおよそ  $x_1$  付近で極値を取ることが期待できる。

たとえば、もし  $f(x)$  が二次関数  $f(x) = ax^2 + bx + c$  であれば

$$g(x) = 2ax + b \quad \text{and} \quad h(x) = 2a \quad (71)$$

であるから、 $x_0$  の値によらず

$$x_1 = -\frac{b}{2a} \quad (72)$$

となるが、これは二次関数の頂点を与える  $x$  座標である。

上記の議論から 1 変数の関数  $f(x)$  の Newton Raphson 法は以下ようになる。

$$x_1 = x_0 - \frac{g(x_0)}{h(x_0)} \quad (73)$$

$$x_2 = x_1 - \frac{g(x_1)}{h(x_1)} \quad (74)$$

$$\vdots \quad \vdots \quad (75)$$

$$x_{\ell+1} = x_\ell - \frac{g(x_\ell)}{h(x_\ell)} \quad (76)$$

なお、これであまりうまくいかない場合には、先に述べたステップサイズ  $\alpha$  を以下のように導入し、Backward Line Search 等を行う必要がある。

$$x_{\ell+1} = x_\ell - \alpha \frac{g(x_\ell)}{h(x_\ell)} \quad (77)$$

### 3.2.2 数値例

$$f(x) = ax^3 + bx^2 + cx + d \quad (78)$$

$$g(x) = 3ax^2 + 2bx + c \quad \text{and} \quad h(x) = 6ax + 2b \quad (79)$$

ただし  $a = 1; b = 0; c = -2; d = 0$  とする。これを  $x_0 = 1.5$  から始めると以下ようになる。

Table1  $f(x) = x^3 - 2x$  の最小化

	xp	g	h	d	f(xp)	x	f(x)
0	NA	NA	NA	NA	NA	1.5000	0.3750
1	1.5000	4.7500	9.0000	0.5278	0.3750	0.9722	-1.0255
2	0.9722	0.8356	5.8333	0.1433	-1.0255	0.8290	-1.0883
3	0.8290	0.0616	4.9738	0.0124	-1.0883	0.8166	-1.0887
4	0.8166	0.0005	4.8995	0.0001	-1.0887	0.8165	-1.0887
5	0.8165	0.0000	4.8990	0.0000	-1.0887	0.8165	-1.0887