一般化線形モデル（推定）

1 一般化線形モデル

一般化線形モデルは，指数分布族に属する分布に従い，かつ以下に示す性質をもつ独立な確率変数 $Y_1, Y_2, \ldots, Y_n$ の集合に対して定義される。

1. 各確率変数 Y_i の分布は，正準形をもち，1 つのパラメータ θ_i（ただし，θ_i はすべて同じ，つまり $\theta_i = \theta$ である必要はない）に依存している。すなわち Y_i の確率（密度）関数は次式で表すことができる。

$$f(y_i; \theta_i) = \exp[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)]$$

2. 確率変数 $Y_1, Y_2, \ldots, Y_n$ の分布は同じ型（例えば，正規分布，二項分布，ポアソン分布など）であり，b, c, d の下の添え字は不要である。このとき，$Y_1, Y_2, \ldots, Y_n$ の同時確率（密度）関数は次式で表すことができる。

$$\begin{aligned} f(y_1, \ldots, y_n; \theta_1, \ldots, \theta_n) &= \prod_{i=1}^{n} \exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)] \\ &= \exp\left[\sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(\theta_i)\right] \end{aligned}$$

それぞれの観測値に対して θ_i が一つずつ存在している可能性があるため，パラメータ θ_i には直接関心ないことが多い。モデルの設定において，通常 $\beta_0, \beta_1, \ldots, \beta_p$ $(p < n)$ という，より少数のパラメータに関心がある。$E[Y_i] = \mu_i$ として定義される μ_i は，$\beta_0, \beta_1, \ldots, \beta_p$ の関数として表されると仮定する。一般化線形モデルにおいて，μ_i の関数 $g(\mu_i)$ について，次のモデルを仮定する。

$$g(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$$

ただし，関数 g は，単調かつ微分可能な関数であり，連結関数（link function）とよばれる。また，$\boldsymbol{x}_i$ は説明変数の $(p+1) \times 1$ ベクトル（共変量や因子の水準に対するダミー変数），$\boldsymbol{\beta}$ は $(p+1) \times 1$ パラメータベクトルである。$\boldsymbol{x}_i$ はデザイン行列 $\boldsymbol{X}$ の i 番目の行である。

$$\boldsymbol{x}_i = \begin{pmatrix} x_{0i} \\ x_{1i} \\ x_{2i} \\ \vdots \\ x_{pi} \end{pmatrix} = \begin{pmatrix} 1 \\ x_{1i} \\ x_{2i} \\ \vdots \\ x_{pi} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

一般化線形モデル（Generalized Linear Model; GLM）は線形モデルの拡張として提案されたものである。線形回帰モデル，対数線形モデル，ロジスティック回帰モデルなどを統一的に表現できるものである。一般化線形モデルは次の 3 つの成分から規定されるモデルである。

1. **ランダム成分（random conponent）**
   説明変数が与えられたときの目的変数は，（正準形をもつ）指数分布族に従うと仮定される。指数分布族には，正規分布，二項分布，ポアソン分布などがある。

2. **系統的成分（systematic component）**
   説明変数は線形的にモデルに関与するとして，線形予測子を説明変数の線形結合として次のように定義する。
   $$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}$$

3. **連結関数（link function）**
   線形予測子は目的変数の平均 $\mu_i = E[Y_i]$ の関数
   $$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}$$
   と仮定される。線形予測子と平均の関係を規定する関数 $g(\cdot)$ は，連結関数とよばれ，通常滑らかで単調性をもつと仮定する。

最も良く知られている一般化線形モデルの特別な場合は，次のモデルである。

$$E[Y_i] = \mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \quad Y_i \sim N(\mu_i, \sigma^2)$$

ここで，$Y_1, Y_2, \ldots, Y_n$ は独立な確率変数である。このモデルの連結関数は，恒等関数 $g(\mu_i) = \mu_i$ である。このモデルは，次のように表すことができる。

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$$

ここで，$p + 1 < n$ として，

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{p1} \\ 1 & x_{12} & x_{22} & \ldots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{pn} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

とする。

正規分布 $N(\mu, \sigma^2)$ の確率密度関数は，次のように表される。

$$\begin{aligned} f(y; \mu, \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}(y - \mu)^2\right] \\ &= \exp\left[-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right] \\ &= \exp\left[\frac{1}{\sigma^2}\left(-\frac{y^2}{2} + y\mu\right) - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right] \end{aligned}$$

まず，μ を関心のある母数，σ^2 を局外母数とした場合を考える。ここで，

$$a(y) = y, \quad b(\mu) = \frac{\mu}{\sigma^2}, \quad c(\mu) = -\frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2), \quad d(y) = -\frac{y^2}{2\sigma^2}$$

または

$$a(y) = y, \quad b(\mu) = \frac{\mu}{\sigma^2}, \quad c(\mu) = -\frac{\mu^2}{2\sigma^2}, \quad d(y) = -\frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)$$

である。したがって，μ を関心のある母数，σ^2 を局外母数とした正規分布は，指数分布族に属している。さらに，$a(y) = y$ であるので，正準形である。

問 3

$Y_1, Y_2, \ldots, Y_n$ は独立な確率変数で，すべての Y_i $(i = 1, \ldots, n)$ に対して，

$$E[Y_i] = \mu_i = \beta_0 + \log(\beta_1 + \beta_2 x_i), \quad Y_i \sim N(\mu_i, \sigma^2)$$

が成り立つとする。これは一般化線形モデルであるか否かを理由も含めて述べよ。

2 推定

最尤法を用いて一般化線形モデルのパラメータの最尤推定値を求める方法を述べる。特別な場合には明示的な式を求めることができるが，多くの場合は数値的な解法が必要となる。それらの方法は，ニュートン・ラプソン法にもとづくことが多い。本節では，その原理を説明する。

2.1 ニュートン・ラプソン法（単変量）

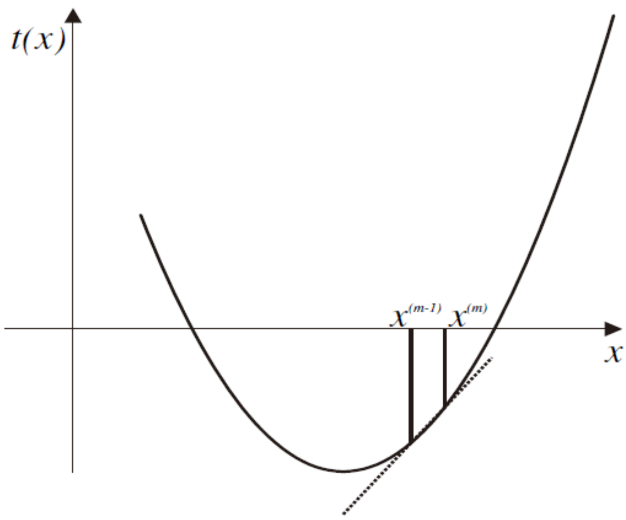

図 1 方程式 $t(x) = 0$ の解を求めるためのニュートン・ラプソン法

図 1 は，ニュートン・ラプソン法の原理を示している。関数 t が x 軸と交わる，すなわち $t(x) = 0$ となる x の値を求ることを考える。値 $x^{(m-1)}$ での t の傾きは，距離 $x^{(m)} - x^{(m-1)}$ が小さければ，近似的に次式によって与えられる。

$$\left[\frac{\partial t}{\partial x}\right]_{x=x^{(m-1)}} = t'(x^{(m-1)}) = \frac{t(x^{(m)}) - t(x^{(m-1)})}{x^{(m)} - x^{(m-1)}}$$

もし，$x^{(m)}$ が $t(x) = 0$ を満たす解ならば，次式を得る。

$$x^{(m)} = x^{(m-1)} - \frac{t(x^{(m-1)})}{t'(x^{(m-1)})} \tag{1}$$

(1) 式が非線形方程式 $t(x) = 0$ を解くためのニュートン・ラプソン法の公式である。初期値 $x^{(1)}$ から始め，(1) 式により逐次近似を行い，収束するまで反復する。

最尤推定の場合について，スコア関数を用いて表すと，(1) 式に対応する推定方程式は次のようになる。

$$\theta^{(m)} = \theta^{(m-1)} - \frac{U^{(m-1)}}{U'^{(m-1)}} \tag{2}$$

最尤推定では，期待値 $E[U']$ により U' を近似することがある。指数分布族に属する分布では,

$$E[U'] = -V[U'] = -\mathfrak{I}$$

が成り立つ。したがって，次式のように，(2) 式の代わりとなる推定方程式が得られる。

$$\theta^{(m)} = \theta^{(m-1)} + \frac{U^{(m-1)}}{\mathfrak{I}}$$

この方法は，**スコア法**と呼ばれる。

期待値 $E[U']$ により U' を近似できることを示す。スコア統計量 U は，対数尤度関数のパラメータ θ に関する導関数である。

$$U(\theta; y) = \frac{\partial l(\theta; y)}{\partial \theta}$$

したがって，次式が成り立つ。

$$U' = \left(\frac{\partial l}{\partial \theta}\right)' = \frac{\partial^2 l}{\partial \theta^2} \Leftrightarrow \frac{1}{n}U' = \frac{1}{n}\left(\frac{\partial^2 l}{\partial \theta^2}\right)$$

ここで，$Y_1, Y_2, \ldots, Y_n$ は独立な確率変数であることから，次式が成り立つ。

$$l = \log \prod_{i=1}^{n} f(y_i) = \sum_{i=1}^{n} \log f(y_i)$$

したがって，次式を得る。

$$\frac{1}{n}U' = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2 \log f(y_i)}{\partial \theta^2}\right)$$

$n \to \infty$ ならば，大数の法則から次式が成り立つ。

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2 \log f(y_i)}{\partial \theta^2}\right) \xrightarrow{P} E\left[\sum_{i=1}^{n}\left(\frac{\partial^2 \log f(y_i)}{\partial \theta^2}\right)\right]$$

さらに,

$$V[U] = -E[U'] = \mathfrak{I} = nE\left[\sum_{i=1}^{n}\left(\frac{\partial^2 \log f(y_i)}{\partial \theta^2}\right)\right]$$

であることから，次式を得る。

$$E\left[\sum_{i=1}^{n}\left(\frac{\partial^2 \log f(y_i)}{\partial \theta^2}\right)\right] \quad \xrightarrow{P} -\frac{1}{n}\mathfrak{I} \Leftrightarrow \frac{1}{n}U' \xrightarrow{P} -\frac{1}{n}\mathfrak{I} \Leftrightarrow U' \xrightarrow{P} -\mathfrak{I} = E[U']$$

したがって，期待値 $E[U']$ により U' を近似できる。

2.2 ニュートン・ラプソン法（多変量）

尤度関数 $L(\boldsymbol{\beta})$ を最大にする $\hat{\boldsymbol{\beta}}$ をニュートン・ラプソン法により求めることを考える。$\boldsymbol{\beta}, \boldsymbol{U}, \boldsymbol{H}$ を次のように定義する。

$$\boldsymbol{U} = \frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0} \\ \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1} \\ \vdots \\ \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_p} \end{pmatrix} = \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_p \end{pmatrix}$$

$$\boldsymbol{H} = \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \begin{pmatrix} \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_0^2} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_p} \\ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1^2} & \cdots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_p \beta_0} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_p \partial \beta_1} & \cdots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_p^2} \end{pmatrix} = \begin{pmatrix} h_{00} & h_{01} & \dots & h_{0p} \\ h_{10} & h_{11} & \dots & h_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ h_{p0} & h_{p1} & \dots & h_{pp} \end{pmatrix}$$

ステップ m で求められた $\boldsymbol{\beta}, \boldsymbol{U}, \boldsymbol{H}$ を $\boldsymbol{\beta}^{(m)}, \boldsymbol{U}^{(m)}, \boldsymbol{H}^{(m)}$ とする。$L(\boldsymbol{\beta})$ を $\boldsymbol{\beta}$ に近い $\boldsymbol{\beta}^{(m)}$ の周りでテイラー展開すると，次式を得る。

$$\begin{aligned} L(\boldsymbol{\beta}) &\approx L(\boldsymbol{\beta}^{(m)}) + \frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)})^\top \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}) \\ &= L(\boldsymbol{\beta}^{(m)}) + (\boldsymbol{U}^{(m)})^\top(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)})^\top \boldsymbol{H}^{(m)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)}) \end{aligned} \tag{3}$$

(3) 式の両辺を $\boldsymbol{\beta}$ で微分すると，次式を得る。

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \approx \frac{\partial L(\boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\beta}} + \frac{\partial (\boldsymbol{U}^{(m)})^\top(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\beta}} + \frac{1}{2}\frac{\partial (\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)})^\top \boldsymbol{H}^{(m)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\beta}} \tag{4}$$

パラメータ $\boldsymbol{\beta}$ の最尤推定量は，$\dfrac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$ を解くことで得られる。

$$\frac{\partial L(\boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial L(\boldsymbol{\beta}^{(m)})}{\partial \beta_0} \\ \frac{\partial L(\boldsymbol{\beta}^{(m)})}{\partial \beta_1} \\ \vdots \\ \frac{\partial L(\boldsymbol{\beta}^{(m)})}{\partial \beta_p} \end{pmatrix} = \mathbf{0}, \tag{5}$$

$$\frac{\partial (\boldsymbol{U}^{(m)})^\top (\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \dfrac{\partial (\boldsymbol{U}^{(m)})^\top (\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})}{\partial \beta_0} \\ \dfrac{\partial (\boldsymbol{U}^{(m)})^\top (\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial (\boldsymbol{U}^{(m)})^\top (\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})}{\partial \beta_p} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial \sum_{i=0}^p u_i^{(m)} (\beta_i - \beta_i^{(m)})}{\partial \beta_0} \\ \dfrac{\partial \sum_{i=0}^p u_i^{(m)} (\beta_i - \beta_i^{(m)})}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial \sum_{i=0}^p u_i^{(m)} (\beta_i - \beta_i^{(m)})}{\partial \beta_p} \end{pmatrix} = \boldsymbol{U}^{(m)} \tag{6}$$

次に，$\dfrac{\partial (\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})^\top \boldsymbol{H}^{(m)} (\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\beta}}$ を整理する。

$$\begin{aligned} &(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})^\top \boldsymbol{H}^{(m)} (\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)}) \\ &= \left(\beta_0 - \beta_0^{(m)}, \beta_1 - \beta_1^{(m)}, \ldots, \beta_p - \beta_p^{(m)}\right) \begin{pmatrix} h_{00}^{(m)} & h_{01}^{(m)} & \ldots & h_{0p}^{(m)} \\ h_{10}^{(m)} & h_{11}^{(m)} & \ldots & h_{1p}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ h_{p0}^{(m)} & h_{p1}^{(m)} & \ldots & h_{pp}^{(m)} \end{pmatrix} \begin{pmatrix} \beta_0 - \beta_0^{(m)} \\ \beta_1 - \beta_1^{(m)} \\ \vdots \\ \beta_p - \beta_p^{(m)} \end{pmatrix} \\ &= \sum_{i=0}^p \sum_{j=0}^p \left(\beta_i - \beta_i^{(m)}\right) h_{ij}^{(m)} \left(\beta_j - \beta_j^{(m)}\right) \end{aligned} \tag{7}$$

(7) 式より，次式を得る。

$$\frac{\partial (\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})^\top \boldsymbol{H}^{(m)} (\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \dfrac{\partial \sum_{i=0}^p \sum_{j=0}^p \left(\beta_i - \beta_i^{(m)}\right) h_{ij}^{(m)} \left(\beta_j - \beta_j^{(m)}\right)}{\partial \beta_0} \\ \dfrac{\partial \sum_{i=0}^p \sum_{j=0}^p \left(\beta_i - \beta_i^{(m)}\right) h_{ij}^{(m)} \left(\beta_j - \beta_j^{(m)}\right)}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial \sum_{i=0}^p \sum_{j=0}^p \left(\beta_i - \beta_i^{(m)}\right) h_{ij}^{(m)} \left(\beta_j - \beta_j^{(m)}\right)}{\partial \beta_p} \end{pmatrix}$$

$$\begin{aligned} &\frac{\partial \sum_{i=0}^p \sum_{j=0}^p \left(\beta_i - \beta_i^{(m)}\right) h_{ij}^{(m)} \left(\beta_j - \beta_j^{(m)}\right)}{\partial \beta_k} \\ &= \frac{\partial \sum_{i=0}^p \left(\beta_i - \beta_i^{(m)}\right) h_{ik}^{(m)} \left(\beta_k - \beta_k^{(m)}\right)}{\partial \beta_k} + \frac{\partial \sum_{i=0}^p \sum_{j\neq k} \left(\beta_i - \beta_i^{(m)}\right) h_{ij}^{(m)} \left(\beta_j - \beta_j^{(m)}\right)}{\partial \beta_k} \\ &= 2\left(\beta_k - \beta_k^{(m)}\right) h_{kk}^{(m)} + \sum_{i\neq k} \left(\beta_i - \beta_i^{(m)}\right) h_{ik}^{(m)} + \sum_{j\neq k} h_{kj}^{(m)} \left(\beta_j - \beta_j^{(m)}\right) \\ &= 2\sum_{i=0}^p \left(\beta_i - \beta_i^{(m)}\right) h_{ik}^{(m)} \quad (\because h_{ik}^{(m)} = h_{ki}^{(m)}) \end{aligned} \tag{8}$$

したがって，(8) 式より，次式を得る。

$$
\begin{aligned}
\frac{1}{2}\frac{\partial(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})^\top \boldsymbol{H}^{(m)}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)})}{\partial\boldsymbol{\beta}} &= \frac{1}{2}\begin{pmatrix} 2\sum_{i=0}^{p}\left(\beta_i-\beta_i^{(m)}\right)h_{i1}^{(m)} \\ 2\sum_{i=0}^{p}\left(\beta_i-\beta_i^{(m)}\right)h_{i2}^{(m)} \\ \vdots \\ 2\sum_{i=0}^{p}\left(\beta_i-\beta_i^{(m)}\right)h_{ip}^{(m)} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=0}^{p}\left(\beta_i-\beta_i^{(m)}\right)h_{i1}^{(m)} \\ \sum_{i=0}^{p}\left(\beta_i-\beta_i^{(m)}\right)h_{i2}^{(m)} \\ \vdots \\ \sum_{i=0}^{p}\left(\beta_i-\beta_i^{(m)}\right)h_{ip}^{(m)} \end{pmatrix} \\
&= \boldsymbol{H}^{(m)}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)}) \qquad (9)
\end{aligned}
$$

(5), (6), (9) 式より，次式を得る。

$$
\begin{aligned}
\frac{\partial L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &\approx \mathbf{0}+\boldsymbol{U}^{(m)}+\boldsymbol{H}^{(m)}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(m)}) \quad (\equiv 0) \\
&\Leftrightarrow \boldsymbol{H}^{(m)}\boldsymbol{\beta} = \boldsymbol{H}^{(m)}\boldsymbol{\beta}^{(m)}-\boldsymbol{U}^{(m)} \\
&\Leftrightarrow \boldsymbol{\beta} = \boldsymbol{\beta}^{(m)}-(\boldsymbol{H}^{(m)})^{-1}\boldsymbol{U}^{(m)} \\
&\Leftrightarrow \boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)}-(\boldsymbol{H}^{(m)})^{-1}\boldsymbol{U}^{(m)} \qquad (10)
\end{aligned}
$$

(10) 式が $\dfrac{\partial L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=\mathbf{0}$ を解くためのニュートン・ラプソン法の公式である。初期値 $\boldsymbol{\beta}^{(1)}$ から始め，(10) 式により逐次近似を行い，収束するまで反復する。

3 一般化線形モデルにおける最尤推定

一般化線形モデルの性質を満たす独立な確率変数 $Y_1, Y_2, \ldots, Y_n$ について考える。つまり，$E[Y_i]=\mu_i$ および $g(\mu_i)=\boldsymbol{x}_i^\top\boldsymbol{\beta}$ という関係式によって Y_i に関係するパラメータ $\boldsymbol{\beta}$ を推定することを考える。

Y_i $(i=1,\ldots,n)$ は正準形をもつ指数分布族に属するので，対数尤度関数は次のようになる。

$$
l_i = y_i b(\theta_i) + c(\theta_i) + d(y_i)
$$

さらに，

$$
\begin{aligned}
E[Y_i] &= \mu_i = -\frac{c'(\theta_i)}{b'(\theta_i)} \\
V[Y_i] &= \frac{b''(\theta_i)c'(\theta_i)-c''(\theta_i)b'(\theta_i)}{(b'(\theta_i))^3} \\
g(\mu_i) &= \boldsymbol{x}_i^\top\boldsymbol{\beta} = \eta_i
\end{aligned}
$$

である。

すべての Y_i に対する，同時対数尤度関数は，次のようになる。

$$
l = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i)
$$

パラメータ β_j の最尤推定量を得るため，微分の連鎖法則に基づく次式を利用する。

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \beta_j} \tag{11}$$

(11) 式の右辺の各項について考える。

$$\begin{aligned}
\frac{\partial l_i}{\partial \theta_i} &= y_i b'(\theta_i) + c'(\theta_i) \\
&= y_i b'(\theta_i) - \mu_i b'(\theta_i) \quad \left(\because \mu_i = -\frac{c'(\theta_i)}{b'(\theta_i)}\right) \\
&= b'(\theta_i)(y_i - \mu_i) \\
\frac{\partial \theta_i}{\partial \mu_i} &= \frac{1}{\left(\dfrac{\partial \theta_i}{\partial \mu_i}\right)} \\
\frac{\partial \mu_i}{\partial \theta_i} &= -\left(\frac{c''(\theta_i)b'(\theta_i) - c'(\theta_i)b''(\theta_i)}{(b'(\theta_i))^2}\right) \\
&= b'(\theta_i) \cdot \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{(b'(\theta_i))^3} \\
&= b'(\theta_i)V[Y_i] \\
\frac{\partial \mu_i}{\partial \beta_j} &= \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ji}
\end{aligned}$$

これらの結果を (11) 式に代入すると，スコア U_j は次式のように表すことができる。

$$\begin{aligned}
U_j &= \sum_{i=1}^{n} \left[b'(\theta_i)(Y_i - \mu_i) \cdot \frac{1}{b'(\theta_i)V[Y_i]} \cdot \frac{\partial \mu_i}{\partial \eta_i} x_{ji} \right] \\
&= \sum_{i=1}^{n} \left[\frac{Y_i - \mu_i}{V[Y_i]} \cdot x_{ji} \cdot \frac{\partial \mu_i}{\partial \eta_i} \right]
\end{aligned}$$

U_j $(j = 0, 1, \ldots, p)$ の分散共分散行列は，情報行列 $\mathfrak{I}$ とよばれる。情報行列 $\mathfrak{I}$ の (k, l) 要素を次のように表す。

$$\mathfrak{I}_{kl} = E[U_k U_l]$$

ここで，$Y_1, Y_2, \ldots, Y_n$ は独立であることから，共分散 $E[(Y_k - \mu_k)(Y_l - \mu_l)] = 0$ $(k \neq l)$ となる。したがって，U_j は次のように表せる。

$$\begin{aligned}
\mathfrak{I}_{kl} &= E[U_k U_l] \\
&= E\left\{ \sum_{s=1}^{n} \left[\frac{Y_s - \mu_s}{V[Y_s]} \cdot x_{ks} \cdot \frac{\partial \mu_s}{\partial \eta_s} \right] \sum_{t=1}^{n} \left[\frac{Y_t - \mu_t}{V[Y_t]} \cdot x_{lt} \cdot \frac{\partial \mu_t}{\partial \eta_t} \right] \right\} \\
&= \sum_{s=1}^{n} \sum_{t=1}^{n} E\left[\frac{Y_s - \mu_s}{V[Y_s]} \frac{Y_t - \mu_t}{V[Y_t]} \cdot x_{ks} x_{lt} \cdot \frac{\partial \mu_s}{\partial \eta_s} \frac{\partial \mu_t}{\partial \eta_t} \right] \\
&= \sum_{s=1}^{n} \left[\frac{E[(Y_s - \mu_s)^2]}{(V[Y_s])^2} \cdot x_{ks} x_{ls} \cdot \left(\frac{\partial \mu_s}{\partial \eta_s} \right)^2 \right] \quad (\because E[(Y_s - \mu_s)(Y_t - \mu_t)] = 0 \ (s \neq t)) \\
&= \sum_{s=1}^{n} \left[\frac{x_{ks} x_{ls}}{V[Y_s]} \cdot \left(\frac{\partial \mu_s}{\partial \eta_s} \right)^2 \right] \quad (\because V[Y_s] = E[(Y_s - \mu_s)^2])
\end{aligned}$$

単一のパラメータの場合のスコア法の推定方程式は，パラメータベクトル $\boldsymbol{\beta}$ に対して，次のように一般化される。

$$\hat{\boldsymbol{\beta}}^{(m)} = \hat{\boldsymbol{\beta}}^{(m-1)} + \left[\boldsymbol{\mathfrak{I}}^{(m-1)}\right]^{-1} \boldsymbol{U}^{(m-1)} \tag{12}$$

ここで，$\hat{\boldsymbol{\beta}}^{(m)}$ は m 回目の反復時におけるパラメータ $\beta_0, \beta_1, \ldots, \beta_p$ の推定値ベクトルである。$\left[\boldsymbol{\mathfrak{I}}^{(m-1)}\right]^{-1}$ は，要素 $\mathfrak{I}_{kl}$ をもつ情報行列 $\boldsymbol{\mathfrak{I}}$ の逆行列であり，$\boldsymbol{U}^{(m-1)}$ は U_j を要素にもつベクトル（ただし，すべて $\hat{\boldsymbol{\beta}}^{(m-1)}$ を用いて計算した値）である。

(12) 式の両辺に $\boldsymbol{\mathfrak{I}}^{(m-1)}$ を左から掛けると，次式を得る。

$$\boldsymbol{\mathfrak{I}}^{(m-1)}\hat{\boldsymbol{\beta}}^{(m)} = \boldsymbol{\mathfrak{I}}^{(m-1)}\hat{\boldsymbol{\beta}}^{(m-1)} + \boldsymbol{U}^{(m-1)} \tag{13}$$

次に，$\mathfrak{I}_{kl} = \displaystyle\sum_{s=1}^{n} \left[\frac{x_{ks}x_{ls}}{V[Y_s]} \cdot \left(\frac{\partial \mu_s}{\partial \eta_s}\right)^2\right]$ の関係式を行列で表すことを考える。

$$\begin{aligned}
\boldsymbol{\mathfrak{I}} &= \begin{pmatrix} \mathfrak{I}_{00} & \mathfrak{I}_{01} & \ldots & \mathfrak{I}_{0p} \\ \mathfrak{I}_{10} & \mathfrak{I}_{11} & \ldots & \mathfrak{I}_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathfrak{I}_{p0} & \mathfrak{I}_{p1} & \ldots & \mathfrak{I}_{pp} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{s=1}^{n}\left[\frac{x_{0s}x_{0s}}{V[Y_s]}\cdot\left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2\right] & \sum_{s=1}^{n}\left[\frac{x_{0s}x_{1s}}{V[Y_s]}\cdot\left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2\right] & \ldots & \sum_{s=1}^{n}\left[\frac{x_{0s}x_{ps}}{V[Y_s]}\cdot\left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2\right] \\
\sum_{s=1}^{n}\left[\frac{x_{1s}x_{0s}}{V[Y_s]}\cdot\left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2\right] & \sum_{s=1}^{n}\left[\frac{x_{1s}x_{1s}}{V[Y_s]}\cdot\left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2\right] & \ldots & \sum_{s=1}^{n}\left[\frac{x_{1s}x_{ps}}{V[Y_s]}\cdot\left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2\right] \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{s=1}^{n}\left[\frac{x_{ps}x_{0s}}{V[Y_s]}\cdot\left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2\right] & \sum_{s=1}^{n}\left[\frac{x_{ps}x_{1s}}{V[Y_s]}\cdot\left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2\right] & \ldots & \sum_{s=1}^{n}\left[\frac{x_{ps}x_{ps}}{V[Y_s]}\cdot\left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2\right] \end{pmatrix} \\
&= \begin{pmatrix} x_{01} & x_{02} & \ldots & x_{0n} \\ x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \ldots & x_{pn} \end{pmatrix}
\begin{pmatrix} \frac{x_{01}}{V[Y_1]}\cdot\left(\frac{\partial\mu_1}{\partial\eta_1}\right)^2 & \frac{x_{11}}{V[Y_1]}\cdot\left(\frac{\partial\mu_1}{\partial\eta_1}\right)^2 & \ldots & \frac{x_{p1}}{V[Y_1]}\cdot\left(\frac{\partial\mu_1}{\partial\eta_1}\right)^2 \\
\frac{x_{02}}{V[Y_2]}\cdot\left(\frac{\partial\mu_2}{\partial\eta_2}\right)^2 & \frac{x_{12}}{V[Y_2]}\cdot\left(\frac{\partial\mu_2}{\partial\eta_2}\right)^2 & \ldots & \frac{x_{p2}}{V[Y_2]}\cdot\left(\frac{\partial\mu_2}{\partial\eta_2}\right)^2 \\
\vdots & \vdots & \ddots & \vdots \\
\frac{x_{0n}}{V[Y_n]}\cdot\left(\frac{\partial\mu_n}{\partial\eta_n}\right)^2 & \frac{x_{1n}}{V[Y_n]}\cdot\left(\frac{\partial\mu_n}{\partial\eta_n}\right)^2 & \ldots & \frac{x_{pn}}{V[Y_n]}\cdot\left(\frac{\partial\mu_n}{\partial\eta_n}\right)^2 \end{pmatrix} \\
&= \boldsymbol{X}^\top \begin{pmatrix} \frac{1}{V[Y_1]}\cdot\left(\frac{\partial\mu_1}{\partial\eta_1}\right)^2 & & & \text{\Large 0} \\ & \frac{1}{V[Y_2]}\cdot\left(\frac{\partial\mu_2}{\partial\eta_2}\right)^2 & & \\ & & \ddots & \\ \text{\Large 0} & & & \frac{1}{V[Y_n]}\cdot\left(\frac{\partial\mu_n}{\partial\eta_n}\right)^2 \end{pmatrix} \boldsymbol{X} \\
&= \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}
\end{aligned}$$

したがって，(13) 式の右辺は次のように表せる。

$$
\begin{aligned}
&\boldsymbol{\mathfrak{I}}^{(m-1)}\hat{\boldsymbol{\beta}}^{(m-1)} + \boldsymbol{U}^{(m-1)} \\
&\quad = \begin{pmatrix} \mathfrak{I}_{00}^{(m-1)} & \mathfrak{I}_{01}^{(m-1)} & \cdots & \mathfrak{I}_{0p}^{(m-1)} \\ \mathfrak{I}_{10}^{(m-1)} & \mathfrak{I}_{11}^{(m-1)} & \cdots & \mathfrak{I}_{1p}^{(m-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathfrak{I}_{p0}^{(m-1)} & \mathfrak{I}_{p1}^{(m-1)} & \cdots & \mathfrak{I}_{pp}^{(m-1)} \end{pmatrix} \begin{pmatrix} \hat{\beta}_0^{(m-1)} \\ \hat{\beta}_1^{(m-1)} \\ \vdots \\ \hat{\beta}_p^{(m-1)} \end{pmatrix} + \begin{pmatrix} U_0^{(m-1)} \\ U_1^{(m-1)} \\ \vdots \\ U_p^{(m-1)} \end{pmatrix} \\
&\quad = \begin{pmatrix} \displaystyle\sum_{t=0}^{p} \mathfrak{I}_{0t}^{(m-1)} \hat{\beta}_t^{(m-1)} + U_0^{(m-1)} \\ \displaystyle\sum_{t=0}^{p} \mathfrak{I}_{1t}^{(m-1)} \hat{\beta}_t^{(m-1)} + U_1^{(m-1)} \\ \vdots \\ \displaystyle\sum_{t=0}^{p} \mathfrak{I}_{pt}^{(m-1)} \hat{\beta}_t^{(m-1)} + U_p^{(m-1)} \end{pmatrix} \\
&\quad = \begin{pmatrix} \displaystyle\sum_{t=0}^{p}\sum_{s=1}^{n} \left[\frac{x_{0s}x_{ts}}{V[Y_s]} \cdot \left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2 \right] \hat{\beta}_t^{(m-1)} + \sum_{i=1}^{n} \left[\frac{Y_i - \mu_i}{V[Y_i]} \cdot x_{0i} \cdot \frac{\partial\mu_i}{\partial\eta_i} \right] \\ \displaystyle\sum_{t=0}^{p}\sum_{s=1}^{n} \left[\frac{x_{1s}x_{ts}}{V[Y_s]} \cdot \left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2 \right] \hat{\beta}_t^{(m-1)} + \sum_{i=1}^{n} \left[\frac{Y_i - \mu_i}{V[Y_i]} \cdot x_{1i} \cdot \frac{\partial\mu_i}{\partial\eta_i} \right] \\ \vdots \\ \displaystyle\sum_{t=0}^{p}\sum_{s=1}^{n} \left[\frac{x_{ps}x_{ts}}{V[Y_s]} \cdot \left(\frac{\partial\mu_s}{\partial\eta_s}\right)^2 \right] \hat{\beta}_t^{(m-1)} + \sum_{i=1}^{n} \left[\frac{Y_i - \mu_i}{V[Y_i]} \cdot x_{pi} \cdot \frac{\partial\mu_i}{\partial\eta_i} \right] \end{pmatrix} \\
&\quad = \begin{pmatrix} \frac{x_{01}}{V[Y_1]} \cdot \left(\frac{\partial\mu_1}{\partial\eta_1}\right)^2 & \frac{x_{02}}{V[Y_2]} \cdot \left(\frac{\partial\mu_2}{\partial\eta_2}\right)^2 & \cdots & \frac{x_{0n}}{V[Y_n]} \cdot \left(\frac{\partial\mu_n}{\partial\eta_n}\right)^2 \\ \frac{x_{11}}{V[Y_1]} \cdot \left(\frac{\partial\mu_1}{\partial\eta_1}\right)^2 & \frac{x_{12}}{V[Y_2]} \cdot \left(\frac{\partial\mu_2}{\partial\eta_2}\right)^2 & \cdots & \frac{x_{1n}}{V[Y_n]} \cdot \left(\frac{\partial\mu_n}{\partial\eta_n}\right)^2 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_{p1}}{V[Y_1]} \cdot \left(\frac{\partial\mu_1}{\partial\eta_1}\right)^2 & \frac{x_{p2}}{V[Y_2]} \cdot \left(\frac{\partial\mu_2}{\partial\eta_2}\right)^2 & \cdots & \frac{x_{pn}}{V[Y_n]} \cdot \left(\frac{\partial\mu_n}{\partial\eta_n}\right)^2 \end{pmatrix} \begin{pmatrix} \displaystyle\sum_{t=0}^{p} x_{t1}\hat{\beta}_t^{(m-1)} + (Y_1 - \mu_1) \cdot \frac{\partial\eta_1}{\partial\mu_1} \\ \displaystyle\sum_{t=0}^{p} x_{t2}\hat{\beta}_t^{(m-1)} + (Y_2 - \mu_2) \cdot \frac{\partial\eta_2}{\partial\mu_2} \\ \vdots \\ \displaystyle\sum_{t=0}^{p} x_{tn}\hat{\beta}_t^{(m-1)} + (Y_n - \mu_n) \cdot \frac{\partial\eta_n}{\partial\mu_n} \end{pmatrix} \\
&\quad = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{z}
\end{aligned}
$$

以上の結果から，(13) 式は次のように表せる。

$$
\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X} \hat{\boldsymbol{\beta}}^{(m)} = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{z} \tag{14}
$$

(14) 式は，$\boldsymbol{z}$ と $\boldsymbol{W}$ が $\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{z}$ に依存するため，反復的に解かなければならないことを除くと，線形モデルに対して重み付き最小二乗法を適用して得られる正規方程式と同じ形をしている。このように，一般化線形モデルに対する最尤推定量は，**反復重み付き最小二乗法**の手順で求められる。多くの統計ソフトウェアでは，一般化線形モデルを当てはめる手順として (14) 式に基づく効率的なアルゴリズムが利用されている。

参考文献

[1] 江金芳．(2016)．一般化線形モデル．朝倉書店.

[2] Dobson, A. J and Barnett, A. G. (2018). *An introduction to generalized linear models, 4th Edition.* Chapman and Hall/CRC.