第5章　指数型分布族と一般化線型モデル

5.1　指数型分布族

Dobson の 3 章を借用して指数分布族と一般化線型モデルを借用．．そして, 岡留の『機械学習』の４章の分類のところを借用する．

$\Theta \subset \mathbb{R}$ を母数空間とし, 母数 $\theta \in \Theta$ をもつ確率分布に従う確率変数 X を考える．

eq:exp-1

定義 5.1. 確率変数 X の p.d.f. または p.m.f.(これを p と記す) の対数をとったものが次の形で記述されるとき, その分布は**指数型分布族**に属すると言われる．

$$\log \mathsf{p}(x|\,\theta) = t(x)b(\theta) + \kappa(\theta) + d(x).$$

ただし $\log 0 = 0$ と約束し, t, b, κ, d は関数で

$$t : \mathbb{X} \longrightarrow \mathbb{R}, \quad b : \Theta \longrightarrow \mathbb{R}, \quad \kappa : \Theta \longrightarrow \mathbb{R}, \quad d : \mathbb{X} \longrightarrow \mathbb{R}$$

とする．ただしこれらの関数形は既知とし, $\mathbb{X} = \{x \in \mathbb{R};\ \mathsf{p}(x|\,\theta) > 0\}$ である．

$t(x) = x$ のとき, その分布は**正準型** (canonical form) であると言われ, $b(\theta)$ は**自然母数** (natural parameter) と呼ばれる．

re:exp-2

注意 5.2. もし関心のある未知母数 θ 以外に他の未知母数があるとき, それらを関数 b, κ を構成する**局外母数** (nuisance parameter) と呼ばれる．
□

ex:exp-3

例 5.3. (Bernoulli 分布) X は $\mathsf{Ber}(\theta)\,(0 < \theta < 1)$ に従うとき, X の p.m.f. p は

$$\mathsf{p}(x|\,\theta) = \theta^x(1-\theta)^{1-x} \qquad (x = 0,\, 1)$$

と書ける．これは

$$\log \mathsf{p}(x|\,\theta) = x\log\theta + (1-x)\log(1-\theta) = x\log\frac{\theta}{1-\theta} + \log(1-\theta)$$

と表現できるので

$$t(x) = x, \quad b(\theta) = \frac{\theta}{1-\theta}, \quad \kappa(\theta) = \log(1-\theta), \quad d(\theta) = 0.$$

□

ex:exp-4

例 5.4. (ポアソン分布) X は $\mathsf{Pois}(\theta)\,(\theta > 0)$ に従うとき, X の p.m.f. p は

$$\mathsf{p}(x|\,\theta) = \frac{\theta^x e^{-\theta}}{x!} \qquad (x = 0, 1, \ldots)$$

と書ける. ただし, $0! = 1$ と形式的に定める. これは

$$\log \mathsf{p}(x|\,\theta) = x \log \theta - \theta - \log x!$$

と表現できるので

$$t(x) = x, \quad b(\theta) = \log \theta, \quad \kappa(\theta) = -\theta, \quad d(x) = -\log x!.$$

□

ex:exp-5

例 5.5. (正規分布) X は $\mathsf{N}(\theta, \sigma^2)\,(\theta \in \mathbb{R}, 0 < \sigma < \infty)$ に従うとき, X の p.d.f. p は

$$\mathsf{p}(x|\,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2}\right] \quad (-\infty < x < \infty)$$

と書かれる. ここで, 関心のある母数は θ で, σ は局外母数で既知とする. このとき,

$$\log \mathsf{p}(x|\,\theta) = -\frac{x^2}{2\sigma^2} + \frac{x\theta}{\sigma^2} - \frac{\theta^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)$$

と表現できるので,

$$t(x) = x, \quad b(\theta) = \frac{\theta}{\sigma^2}, \quad \kappa(\theta) = -\frac{\theta^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2), \quad d(x) = -\frac{x^2}{2\sigma^2}.$$

□

5.2 指数型分布族の性質

sec:exp-2

以下では連続型分布の場合で議論を進める. 離散型分布の場合には積分記号を和の記号に書きかえればよい. さらに積分記号と微分記号の交換が可能であることが保証されていると仮定[1]して議論を進める.

まず $\mathsf{p}(x|\,\theta)$ は p.d.f. であることから

$$\int_{\mathcal{X}} \mathsf{p}(x|\,\theta)\,dx = 1. \tag{5.1}$$

eq:exp-1

[1]実は, 指数型分布族の場合には積分記号と微分記号の交換が保証されることを示すことができる. Lehmann and Romano (2022, pp.52-53) を参照のこと.

ただし $\mathcal{X} = \{x \in \mathbb{R};\ \mathsf{p}(x|\,\theta) > 0\}$ である. θ に関して (5.1) の両辺を微分すれば,

$$0 = \frac{d}{d\theta}1 = \frac{d}{d\theta}\int_{\mathbb{X}} \mathsf{p}(x|\,\theta)\,dx = \int_{\mathbb{X}} \frac{d}{d\theta}\mathsf{p}(x|\,\theta)\,dx$$

から

$$\int_{\mathbb{X}} \frac{d}{d\theta}\mathsf{p}(x|\,\theta)\,dx = 0 \tag{5.2}$$

ex:exp-2

を得る. 同様に

$$\int_{\mathbb{X}} \frac{d^2\mathsf{p}(x|\,\theta)}{d\theta^2}\,dx = 0. \tag{5.3}$$

eq:exp-3

指数型分布族に属する分布の p.d.f. を (5.1) に代入すると

$$\begin{aligned}\int_{\mathbb{X}} \frac{d}{d\theta}\mathsf{p}(x|\,\theta)\,dx &= \int_{\mathbb{X}} \left(\frac{d}{d\theta}\log \mathsf{p}(x|\,\theta)\right)\mathsf{p}(x|\,\theta)\,dx \\ &= \int_{\mathbb{X}} \left\{t(x)\dot{b}(\theta) + \dot{\kappa}(\theta)\right\}\mathsf{p}(x|\,\theta)\,dx = 0.\end{aligned}$$

ただし

$$\dot{b}(\theta) = \frac{db(\theta)}{d\theta}, \quad \dot{\kappa}(\theta) = \frac{d\kappa(\theta)}{d\theta}.$$

期待値の定義から

$$\int_{\mathbb{X}} t(x)\dot{b}(\theta)\mathsf{p}(x|\,\theta)\,dx = \dot{b}(\theta)\mathsf{E}[a(X)], \quad \int_{\mathbb{X}} \dot{\kappa}(\theta)\mathsf{p}(x|\,\theta)\,dx = \dot{\kappa}(\theta)$$

となるので,

$$\dot{b}(\theta)\mathsf{E}[a(X)] + \dot{\kappa}(\theta) = 0. \tag{5.4}$$

eq:exp-4

よって $\dot{b}(\theta) \neq 0$ のとき

$$\mathsf{E}[t(X)] = -\frac{\dot{\kappa}(\theta)}{\dot{b}(\theta)}. \tag{5.5}$$

eq:exp-5

同様に

$$\begin{aligned}\frac{d^2}{d\theta^2}\mathsf{p}(x|\,\theta) &= \frac{d}{d\theta}\left[\left(\frac{d}{d\theta}\log \mathsf{p}(x|\,\theta)\right)\mathsf{p}(x|\,\theta)\right] \\ &= \left(\frac{d^2}{d\theta^2}\log \mathsf{p}(x|\,\theta)\right)\mathsf{p}(x|\,\theta) + \left(\frac{d}{d\theta}\log \mathsf{p}(x|\,\theta)\right)^2\mathsf{p}(x|\,\theta) \\ &= \left\{a(x)\ddot{b}(\theta) + \ddot{c}(\theta)\right\}\mathsf{p}(x|\,\theta) + \left\{t(x)\dot{b}(\theta) + \dot{\kappa}(\theta)\right\}^2\mathsf{p}(x|\,\theta).\end{aligned} \tag{5.6}$$

eq:exp-6

ただし

$$\ddot{b}(\theta) = \frac{d^2 b(\theta)}{d\theta^2}, \quad \ddot{c}(\theta) = \frac{d^2 c(\theta)}{d\theta^2}.$$

eq:exp-6 (5.6) の最右辺の第 2 項に eq:exp-5 (5.5) を代入すると

$$\begin{aligned}\big\{t(x)\dot{b}(\theta) + \dot{\kappa}(\theta)\big\}\mathsf{p}(x|\,\theta) &= \{\dot{b}(\theta)\}^2\left\{t(x) + \frac{\dot{\kappa}(\theta)}{\dot{b}(\theta)}\right\}^2 \mathsf{p}(x|\,\theta) \\ &= \{\dot{b}(\theta)\}^2\big\{t(x) - \mathsf{E}[t(X)]\big\}^2 \mathsf{p}(x|\,\theta)\end{aligned}$$

と表され

$$\int_{\mathbb{X}} \{t(x) - \mathsf{E}[t(X)]\}^2 \mathsf{p}(x|\,\theta)\,dx = \mathsf{Var}[t(X)]. \tag{5.7}$$

eq:exp-7

一方, eq:exp-3 (5.3) と eq:exp-7 (5.7) から eq:exp-6 (5.6) は以下のように書き直せる.

$$\begin{aligned}0 &= \int_{\mathbb{X}} \frac{d^2}{d\theta^2}\mathsf{p}(x|\,\theta)\,dx = \int_{\mathbb{X}} \big\{t(x)\ddot{b}(\theta) + \ddot{\kappa}(\theta)\big\}\mathsf{p}(x|\,\theta)\,dx + \{\dot{b}(\theta)\}^2 \mathsf{Var}[t(X)] \\ &= \ddot{b}(\theta)\mathsf{E}[t(X)] + \ddot{\kappa}(\theta) + \{\dot{b}(\theta)\}^2 \mathsf{Var}[t(X)].\end{aligned}$$

よって $\dot{b}(\theta) \neq 0$ のとき

$$\mathsf{Var}[t(X)] = \frac{\ddot{b}(\theta)\dot{\kappa}(\theta) - \ddot{\kappa}(\theta)\dot{b}(\theta)}{\{\dot{b}(\theta)\}^3}. \tag{5.8}$$

eq:exp-7a

次に $X = x$ が観測されたときの対数尤度関数

$$\ell(\theta|\,x) = \log \mathsf{p}(x|\,\theta)$$

の導関数の x に確率変数 X を形式的に代入して確率変数にしたもの $\ell(\theta|\,X)$ (これも対数尤度関数と乱暴に呼ぶことにする) の期待値と分散を求める.

$$S(\theta|\,x) := \frac{\partial \ell(\theta|\,x)}{\partial\theta} \quad = \frac{\partial}{\partial\theta}\{t(x)b(\theta) + \kappa(\theta) + d(x)\}\} = t(x)\dot{b}(\theta) + \dot{\kappa}(\theta).$$

うえの式の x に形式的に X を代入したもの $S(\theta|\,X)$ を**スコア関数**という. すると eq:exp-5 (5.5) から

$$\begin{aligned}\mathsf{E}[S(\theta|\,X)] &= \mathsf{E}[t(X)\dot{b}(\theta) + \dot{\kappa}(\theta)] = \dot{b}(\theta)\mathsf{E}[t(X)] + \dot{\kappa}(\theta) \\ &= \dot{b}\left[-\frac{\dot{\kappa}(\theta)}{\dot{b}(\theta)}\right] + \dot{\kappa}(\theta) = 0.\end{aligned} \tag{5.9}$$

eq:exp-8

$S(\theta|\,X)$ の分散を Fisher **情報量**といい, $I(\theta)$ と記すことにする. このとき

$$I(\theta) = \mathsf{Var}[S(\theta|\,X)] = \mathsf{Var}[a(X)\dot{b}(\theta) + \dot{\kappa}(\theta)] = \{\dot{b}(\theta)\}^2 \mathsf{Var}[t(X)].$$

eq:exp-7
(5.7) から

$$\mathsf{Var}[S(\theta|\,X)] = \{\dot{b}(\theta)\}^2 \mathsf{Var}[t(X)] = \frac{\ddot{b}(\theta)\dot{\kappa}(\theta)}{\dot{b}(\theta)} - \ddot{\kappa}(\theta).$$

さらに

$$\dot{S}(\theta|\,x) := \frac{dS(\theta|\,x)}{d\theta} = t(x)\ddot{b}(\theta) + \ddot{\kappa}(\theta)$$

eq:exp-6
と (5.6) から

$$\begin{aligned}\mathsf{E}[\dot{S}(\theta|\,X)] &= \ddot{b}(\theta)\mathsf{E}[t(X)] + \ddot{\kappa}(\theta) = \ddot{b}(\theta)\left[-\frac{\dot{\kappa}(\theta)}{\dot{b}(\theta)}\right] + \ddot{\kappa}(\theta) \\ &= -\frac{\ddot{b}(\theta)\dot{c}(\theta)}{\dot{b}(\theta)} + \ddot{c}(\theta) = -I(\theta) = -\mathsf{Var}[S(\theta|\,X)].\end{aligned}$$

5.3 一般化線型モデル (Generalized linear models(GLM))

GLM は独立な確率変数列 $Y_1, Y_2, \ldots, Y_n$ に対する統計的モデルで以下で定める性質をみたす.

- 各 $Y_j\,(j = 1, 2, \ldots, n)$ の分布は同じ指数型分布族に属する. Y_j の p.d.f. または p.m.f. $\mathsf{p}_Y(y_j|\,\theta_j)$ は母数 $\theta_j \in \Theta \subset \mathbb{R}$ で添え字づけられ

$$\mathsf{p}_Y(y_j|\,\theta_j) = \exp\bigl\{y_j b(\theta_j) + \kappa(\theta_j) + d(y_j)\bigr\} \tag{5.10}$$

eq:glm-1

  で与えられる.

- 母数 θ_j は直接興味ある対象ではなく, 母数 $\beta_1, \beta_2, \ldots, \beta_d\,(d < n)$ に興味がある. ここで

$$\mu_j = \mathsf{E}[Y_j]$$

  と書いたとき

$$g(\mu_j) = \boldsymbol{x}_j^\top \boldsymbol{\beta}$$

  と表現される. ただし $g : \mathbb{R} \longrightarrow \mathbb{R}$ は狭義単調関数で連続微分可能である. したがって μ_j は θ_j の関数で eq:exp-5 (5.5) から

$$\mu_j = -\frac{\dot{\kappa}(\theta_j)}{\dot{b}(\theta_j)}$$

と書ける．さらに $\boldsymbol{x}_j$ は非ランダムな説明変数（既知のもの）で $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_d)^\top$ は未知の回帰係数ベクトルで

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{bmatrix} \in \mathbb{R}^d, \quad \boldsymbol{x}_j = \begin{bmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jd} \end{bmatrix} \in \mathbb{R}^d,$$

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1d} \\ x_{21} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{nd} \end{bmatrix}$$

である．

したがって $(Y_1, Y_2, \ldots, Y_n)$ の同時 p.d.f. または p.m.f. $\mathsf{p}(y_1, y_2, \ldots, y_n | \theta_1, \theta_2, \ldots, \theta_n)$ は

$$\begin{aligned} \mathsf{p}(y_1, y_2, \ldots, y_n | \theta_1, \theta_2, \ldots, \theta_n) &= \prod_{j=1}^n \exp\{y_j b(\theta_j) + \kappa(\theta_j) + d(y_j)\} \\ &= \exp\left\{\sum_{j=1}^n y_j b(\theta_j) + \sum_{j=1}^n \kappa(\theta_j) + \sum_{j=1}^n d(y_j)\right\} \end{aligned} \tag{5.11}$$

eq:glm-2

となる．

ex:glm-1

例 5.6. (ロジスティック回帰モデル)

書くこと．

□

5.4 GLM の最尤推定法

$Y_1, Y_2, \ldots, Y_n$ は (5.10) で与えられる GLM に従うとする．すると $Y_j = y_j\ (j = 1, 2, \ldots, n)$ を観測したときの対数尤度関数 $\ell_j(\theta_j)$ は

eq:glm-1

$$\ell_j(\theta_j) = y_j b(\theta_j) + \kappa(\theta_j) + d(y_j)$$

と書くことができる．さらに (5.5) と (5.8) から

$$\mathsf{E}[Y_j] = \mu_j = -\frac{\dot{\kappa}(\theta_j)}{\dot{b}(\theta_j)} \tag{5.12}$$

eq:glm-3

$$\mathsf{Var}[Y_j] = \frac{\ddot{b}(\theta_j)\dot{\kappa}(\theta_j) - \ddot{\kappa}(\theta_j)\dot{b}(\theta_j)}{\{\dot{b}(\theta_j)\}^3} \tag{5.13}$$

eq:glm-4

$$g(\mu_j) = \boldsymbol{x}_j^\top \boldsymbol{\beta} =: \eta_j, \quad \boldsymbol{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jd})^\top \tag{5.14}$$

eq:glm-5

である．$(Y_1, Y_2, \ldots, Y_n)^\top$ の対数尤度関数 ℓ は

$$\ell = \sum_{j=1}^n \ell_j(\theta_j) = \sum_{j=1}^n y_j b(\theta_j) + \sum_{j=1}^n \kappa(\theta_j) + \sum_{j=1}^n d(y_j)$$

となる．$\boldsymbol{\beta}$ の最尤推定値を求めるために ℓ を $\boldsymbol{\beta}$ に関して偏微分をする．
連鎖偏微分律から

$$\frac{\partial \ell}{\partial \beta_k} = \sum_{j=1}^n \frac{\partial \ell_j}{\partial \beta_k} = \sum_{j=1}^n \frac{\partial \ell_j}{\partial \theta_j} \cdot \frac{\partial \theta_j}{\partial \mu_j} \cdot \frac{\partial \mu_j}{\partial \beta_k} \quad (k = 1, 2, \ldots, d) \tag{5.15}$$

eq:glm-6

と書ける．最右辺の各項を計算して行く．(5.12) に注意すると

$$\frac{\partial \ell_j}{\partial \theta_j} = y_j \dot{b}(\theta_j) + \dot{\kappa}(\theta_j) = \dot{b}(\theta_j)\left\{ y_j + \frac{\dot{\kappa}(\theta_j)}{\dot{b}(\theta_j)} \right\} = \dot{b}(\theta_j)\{y_j - \mu_j\} \tag{5.16}$$

eq:glm-7

を得る．次に

$$\frac{\partial \theta_j}{\partial \mu_j} = 1 \Big/ \frac{\partial \mu_j}{\partial \theta_j}$$

であることと (5.12) に注意すると

$$\begin{aligned}\frac{\partial \mu_j}{\partial \theta_j} &= -\frac{\partial}{\partial \theta_j}\left(\frac{\dot{\kappa}(\theta_j)}{\dot{b}(\theta_j)}\right) = -\frac{\ddot{\kappa}(\theta_j)}{\dot{b}(\theta_j)} + \frac{\ddot{b}(\theta_j)\dot{\kappa}(\theta_j)}{\{\dot{b}(\theta_j)\}^2} \\ &= \dot{b}(\theta_j)\left\{\frac{\ddot{b}(\theta_j)\dot{\kappa}(\theta_j) - \dot{b}(\theta_j)\ddot{\kappa}(\theta_j)}{\{\dot{b}(\theta_j)\}^3}\right\} = \dot{b}(\theta_j)\mathsf{Var}[Y_j]\end{aligned} \tag{5.17}$$

eq:glm-8

を得る．最後の等号は (5.13) からわかる．最後に

$$\frac{\partial \mu_j}{\partial \beta_k} = \frac{\partial \mu_j}{\partial \eta_j} \cdot \frac{\partial \eta_j}{\partial \beta_k} = \frac{\partial \mu_j}{\partial \eta_j} x_{jk} \tag{5.18}$$

eq:glm-9

を得る．(5.16) – (5.18) を (5.15) に代入することでスコア関数は

$$S_k := \sum_{j=1}^n \frac{(Y_j - \mu_j)}{\mathsf{Var}[Y_j]} x_{jk} \left(\frac{\partial \mu_j}{\partial \eta_j}\right) \qquad (k = 1, 2, \ldots, d) \tag{5.19}$$

eq:glm-10

となることがわかる．

$S_1, S_2, \dots, S_d$ の共分散行列を $J = (J_{k\ell})$ は

$$J_{k\ell} = \mathsf{E}[S_k S_\ell] \qquad (k,\, \ell = 1,\, 2,\, \dots,\, d)$$

と書くことができる．すると (5.19) から eq:glm-10

$$\begin{aligned} J_{k\ell} &= \mathsf{E}\left[\sum_{i=1}^{n}\left\{\frac{(Y_i-\mu_i)}{\mathsf{Var}[Y_i]}x_{ik}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\right\}\sum_{j=1}^{n}\left\{\frac{(Y_j-\mu_j)}{\mathsf{Var}[Y_j]}x_{j\ell}\left(\frac{\partial\mu_j}{\partial\eta_j}\right)\right\}\right] \\ &= \sum_{i=1}^{n}\frac{\mathsf{E}[(Y_i-\mu_i)^2]}{\{\mathsf{Var}[Y_i]\}^2}x_{ik}x_{i\ell}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 \end{aligned}$$

を得る．最後から 2 番目の等号は $\mathsf{E}[(Y_i-\mu_i)(Y_j-\mu_j)] = 0\,(i \neq j)$ からわかる．結局

$$J_{k\ell} = \sum_{i=1}^{n}\frac{x_{ik}x_{i\ell}}{\mathsf{Var}[Y_i]}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 \qquad (k,\, \ell = 1,\, 2,\, \dots,\, d) \tag{5.20}$$

eq:glm-11

がわかる．

$m = 1, 2, \dots$ とする．$\widehat{\boldsymbol{\beta}}^{(0)}$ を $\boldsymbol{\beta}$ の初期推定値とし，繰り返し計算で得られる m 回目の $\boldsymbol{\beta}$ の推定値を $\widehat{\boldsymbol{\beta}}^{(m)}$ と書くことにする．$J^{(m)} = (J_{k\ell}^{(m)})$ の $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{(m)}$ での値とし，$\boldsymbol{S}^{(m)} = (S_1^{(m)}, S_2^{(m)}, \dots, S_d^{(m)})^\top$ を S_k の $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{(m)}$ での値とする．したがって (3.4) から eq:fisher-1

$$\widehat{\boldsymbol{\beta}}^{(m)} = \widehat{\boldsymbol{\beta}}^{(m-1)} + \left[\mathbf{J}^{(m-1)}\right]^{-1}\mathbf{S}^{(m-1)} \tag{5.21}$$

eq:glm-12

を得る．上式の両辺に $\mathbf{J}^{(m-1)}$ を掛けると

$$\mathbf{J}^{(m-1)}\widehat{\boldsymbol{\beta}}^{(m)} = \mathbf{J}^{(m-1)}\widehat{\boldsymbol{\beta}}^{(m-1)} + \mathbf{S}^{(m-1)} \tag{5.22}$$

eq:glm-13

を得る．

いま

$$\mathbf{W} = (W_{ij})_{i,\,j=1,\,2,\,\dots,\,n}, \qquad W_{ij} = \begin{cases} \dfrac{1}{\mathsf{Var}[Y_i]}\left(\frac{\partial\mu_i}{\partial\eta_i}\right) & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{5.23}$$

eq:glm-13a

とおき，$\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{(m)}$ での値を $\mathbf{W}^{(m)}$ と書く．すると

$$\mathbf{J}^{(m)} = \mathbf{X}^\top\mathbf{W}^{(m)}\mathbf{X}$$

と書ける．(5.22) eq:glm-13 の右辺のベクトルの第 j 成分 $(j = 1, 2, \dots, d)$ は

$$\sum_{k=1}^{d}\sum_{i=1}^{n}\frac{x_{ij}x_{ik}}{\mathsf{Var}[Y_i]}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\widehat{\beta}_k^{(m-1)} + \sum_{i=1}^{n}\frac{(Y_i-\mu_i)x_{ij}}{\mathsf{Var}[Y_i]}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)$$

と書ける．ただし $\mathsf{Var}[Y_i]$, μ_i, $\frac{\partial \mu_i}{\partial \eta_i}$ も $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{(m-1)}$ での値であることに注意せよ．さらに

$$\begin{aligned} \boldsymbol{z}^{(m)} &= (z_1^{(m)},\, z_2^{(m)},\, \ldots,\, z_d^{(m)})^\top, \\ z_i^{(m)} &= \sum_{k=1}^{d} x_{ik}\widehat{\beta}_k^{(m)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right) + (Y_i - \mu_i) \end{aligned} \tag{5.24}$$

eq:glm-14

とおくと (5.22) の右辺は

$$\mathbf{X}^\top \mathbf{W}^{(m-1)} \widehat{\boldsymbol{z}}^{(m-1)}$$

と書けるので

$$\mathbf{X}^\top \mathbf{W}^{(m-1)} \mathbf{X} \widehat{\boldsymbol{\beta}}^{(m)} = \mathbf{X}^\top \mathbf{W}^{(m-1)} \boldsymbol{z}^{(m-1)}$$

を得る．なぜならば $\mathbf{X}^\top \mathbf{W}^{(m-1)} \boldsymbol{z}^{(m-1)}$ は第 j 成分 を $\{\mathbf{X}^\top \mathbf{W}^{(m-1)} \boldsymbol{z}^{(m-1)}\}_j$ $(j = 1, 2, \ldots, d)$ と書くと

$$\begin{aligned} \{\mathbf{X}^\top \mathbf{W}^{(m-1)} \boldsymbol{z}^{(m-1)}\}_j &= \sum_{i=1}^{n} x_{ij} W_{ii}^{(m-1)} \widehat{z}_i \\ &= \sum_{i=1}^{n} x_{ij} \frac{1}{\mathsf{Var}[Y_i]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\left\{\sum_{k=1}^{d} x_{ik}\widehat{\beta}_k^{(m)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right) + (Y_i - \mu_i)\right\} \\ &\qquad \left(\because \text{(5.23) と (5.24) を代入}\right) \\ &= \sum_{k=1}^{d}\sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{\mathsf{Var}[Y_i]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 + \sum_{i=1}^{n} \frac{(Y_i - \mu_i)x_{ij}}{\mathsf{Var}[Y_i]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right) \end{aligned}$$

からわかる．

例 5.7. (Poissson 回帰)

y_j	2	3	6	7	8	9	10	11	12
x_j	-1	-1	0	0	0	0	1	1	1

応答変数 Y_j は Poisson 分布に従う確率変数とする．すると

$$\mu_j = \mathsf{E}[Y_j] = \mathsf{Var}[Y_j] \qquad (j = 1,\, 2,\, \ldots,\, n;\, n = 9) \tag{5.25}$$

eq:glm-15

となる．μ_j と x_j の間に線型関係を仮定する．

$$\begin{aligned} &\mathsf{E}[Y_j] = \mu_j = \beta_1 + \beta_2 x_j = \boldsymbol{x}_j^\top \boldsymbol{\beta}, \\ &\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad \boldsymbol{x}_j = \begin{bmatrix} 1 \\ x_j \end{bmatrix} \quad (j = 1,\, 2,\, \ldots,\, n) \end{aligned}$$

である．したがってリンク関数 g は恒等関数である．

$$g(\mu_j) = \mu_j = \boldsymbol{x}_j^\top \boldsymbol{\beta} =: \eta_j.$$

これより

$$\frac{\partial \mu_j}{\partial \eta_j} = 1$$

eq:glm-13 eq:glm-15
となる．(5.23) と (5.25) から

$$W_{jj} = \frac{1}{\mathsf{Var}[Y_j]} = \frac{1}{\beta_1 + \beta_2 x_j} \quad (j = 1, 2, \ldots, n)$$

eq:glm-14
となる．$\mu_j = \boldsymbol{x}_j^\top \widehat{\boldsymbol{\beta}} = \widehat{\beta}_1 + \widehat{\beta}_2 x_j$ を (5.24) に代入すると

$$z_j = \widehat{\beta}_1^{(m-1)} + \widehat{\beta}_2^{(m-1)} x_j + y_j - \widehat{\beta}_1^{(m-1)} - \widehat{\beta}_2^{(m-1)} x_j = y_j$$

となる．また

$$\mathbf{J} = \mathbf{X}^\top \mathbf{W}^{(m-1)} \mathbf{X} = \begin{bmatrix} \sum_{j=1}^n \frac{1}{\widehat{\beta}_1^{(m-1)} + \widehat{\beta}_2^{(m-1)} x_j} & \sum_{j=1}^n \frac{x_j}{\widehat{\beta}_1^{(m-1)} + \widehat{\beta}_2^{(m-1)} x_j} \\ \sum_{j=1}^n \frac{x_j}{\widehat{\beta}_1^{(m-1)} + \widehat{\beta}_2^{(m-1)} x_j} & \sum_{j=1}^n \frac{x_j^2}{\widehat{\beta}_1^{(m-1)} + \widehat{\beta}_2^{(m-1)} x_j} \end{bmatrix}$$

$$\mathbf{X}^\top \mathbf{W}^{(m-1)} \boldsymbol{z} = \begin{bmatrix} \sum_{j=1}^n \frac{y_j}{\widehat{\beta}_1^{(m-1)} + \widehat{\beta}_2^{(m-1)} x_j} \\ \sum_{j=1}^n \frac{x_j y_j}{\widehat{\beta}_1^{(m-1)} + \widehat{\beta}_2^{(m-1)} x_j} \end{bmatrix}$$

となる．よって

$$\mathbf{X}^\top \mathbf{W}^{(m-1)} \mathbf{X} \widehat{\boldsymbol{\beta}}^{(m)} = \mathbf{X}^\top \mathbf{W}^{(m-1)} \boldsymbol{y}$$

を得る．

$\widehat{\beta}_1^{(0)} = 7,\ \widehat{\beta}_2^{(0)} = 5$ とする．

$$\boldsymbol{y} = \boldsymbol{z} = \begin{bmatrix} 2 \\ 3 \\ \vdots \\ 12 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_9^\top \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}$$

なので

$$\mathbf{X}^\top \mathbf{W}^{(0)} \mathbf{X} = \begin{bmatrix} 1.821429 & -0.75 \\ -0.75 & 1.25 \end{bmatrix}, \quad \mathbf{X}^\top \mathbf{W}^{(0)} \boldsymbol{y} = \begin{bmatrix} 9.869048 \\ 0.583333 \end{bmatrix}$$

なので

$$\begin{aligned} \widehat{\boldsymbol{\beta}}^{(1)} &= \left[\mathbf{X}^\top \mathbf{W}^{(0)} \mathbf{X}\right]^{-1} \mathbf{X}^\top \mathbf{W}^{(0)} \boldsymbol{y} \\ &= \begin{bmatrix} 0.729167 & 0.4375 \\ 0.4375 & 1.0625 \end{bmatrix} \begin{bmatrix} 9.869048 \\ 0.583333 \end{bmatrix} = \begin{bmatrix} 7.4514 \\ 4.9375 \end{bmatrix} \end{aligned}$$

のように計算される. 最尤推定値は

$$\widehat{\beta}_1 = 7.45163, \quad \widehat{\beta}_2 = 4.93530$$

となり, この値における $\mathbf{J}$ の逆行列の値は

$$\mathbf{J}^{-1} = \begin{bmatrix} 0.7817 & 0.4166 \\ 0.4166 & 1.1863 \end{bmatrix}$$

となる. 繰り返し計算の結果は下記のようになった.

m	0	1	2	3
$\widehat{\beta}_1^{(m)}$	7	7.45139	7.45163	7.45163
$\widehat{\beta}_2^{(m)}$	5	4.93750	4.93531	4.93530

□