10 一般化線型モデル

一般化線型モデルは，9 章の分散分析や 1 章で紹介した寿命調査で用いられた Poisson 回帰を含む回帰モデルの総称である．この章では，一般化線型モデルを一般的な表現で与える．このモデルはパラメトリック分布の一種であるため，大標本では最尤法が適用できる．一般化線型モデルのパラメータ推定量は，スコア方程式を数値的に解くことで得られる．検定や信頼区間は 6 章で述べたように最尤法によって構成することができる．

キーワード 赤池情報量規準（AIC），Kullback–Leibler 情報量，指数型分布族，自然パラメータ，十分統計量，スコア方程式，正準型，正準リンク関数，多重共線性，デザイン行列，分散拡大因子，リンク関数

事　例 6 都市研究

10.1 指数型分布族

■ 10.1.1 定　義

確率変数 Y が，単一のパラメータ θ によって規定される確率分布に従い，確率密度関数または確率関数が

$$p(y;\theta) = \exp\left[a(y)b(\theta) + c(\theta) + d(y)\right]$$

という形式で表されるとき，この分布のクラスを指数型分布族（exponential family）という．関数 $a(x)$，$b(x)$，$c(x)$，$d(x)$ の選び方によって，正規分布や 2 項分布など，指数型分布族のどれかがが決まる．別の言い方をすれば，確率分布によっては扱いづらいものもあるから，指数型分布族だけに注目することで，一般化線型モデルへ拡張しやすくした，といった方がわかりやすいかもしれない．

特に $a(y) = y$ のとき，この分布は正準（canonical）と呼ばれる．正規分布，

2 項分布，Poisson 分布はすべて正準な指数型分布族である．また，$b(\theta)$ のことを自然パラメータ（natural parameter）という．

指数型分布族にはいくつか便利な特徴がある．まず，$a(y)$ の期待値と分散は，$b(\theta)$ と $c(\theta)$ を用いて以下のように書ける．

$$\mathrm{E}[a(Y)] = -\frac{c'(\theta)}{b'(\theta)}$$

$$\mathrm{Var}[a(Y)] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3}$$

また，指数型分布族の対数尤度関数は

$$l(\theta) = a(y)b(\theta) + c(\theta) + d(y)$$

だから，スコア関数はこれを微分して

$$U(\theta) = a(y)b'(\theta) + c'(\theta)$$

となる．また，定理 5-3 より，スコア関数の期待値は

$$\mathrm{E}[U(\theta)] = 0$$

であり，Fisher 情報量はスコア関数の分散であるから

$$I(\theta) = \mathrm{Var}[U(\theta)] = [b'(\theta)]^2\mathrm{Var}[a(Y)]$$

となる．

■ 10.1.2 例：正規分布

正規分布の確率密度関数は，次のように変形できる．

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

ここで

$$b(\mu) = \frac{\mu}{\sigma^2}$$

$$c(\mu) = -\frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)$$

$$d(y) = -\frac{y^2}{2\sigma^2}$$

とおけば，明らかに正準形の指数型分布族であることがわかる．さらに，簡単

な計算から

$$\mathrm{E}(Y) = -\frac{c'(\theta)}{b'(\theta)} = \mu$$

$$\mathrm{Var}(Y) = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3} = \sigma^2$$

であることも確認できる．

10.2　一般化線型モデル

■ 10.2.1　モデルの構造

9 章までは独立同一な指数型分布族の最尤法について述べてきた．それでは，この結果を平均が共通という仮定が満たされない状況に適用するには，どうしたらよいだろうか．一般化線型モデルを提案した Nelder and Wedderburn（1972）のアイデアは，以下のようなものである．まず，Y_i の分布が正準つまり $a(y) = y$ で，θ_i という単一のパラメータによって決まる．さらにそれ以外の関数は，i を通じて共通であると仮定する．

$$p(y_i;\theta_i) = \exp\left[y_i b(\theta_i) + c(\theta_i) + d(y_i)\right]$$

このとき，Y_i（$i = 1,\ldots,N$）の同時分布は

$$\begin{aligned} p(y_1,...,y_N) &= \prod_{i=1}^{N} \exp\left[y_i b(\theta_i) + c(\theta_i) + d(y_i)\right] \\ &= \exp\left[\sum_{i=1}^{N} y_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)\right] \end{aligned}$$

と書くことができる．ただし，個人レベルの N 個のパラメータ θ_i をすべて推定したいわけではない．関心があるのは，Y_i の平均とデザイン行列 $\boldsymbol{X}_i$ の関係を表す

$$g[\mathrm{E}(Y_i|\boldsymbol{X}_i)] = \boldsymbol{X}_i\boldsymbol{\beta}$$

である．このような構造を持つ確率分布のことを一般化線型モデル（generalized linear model）という．このとき個人レベルのパラメータ θ_i は，p 次元のパラメータ $\boldsymbol{\beta}$ によって表される．ただし，パラメータの数は $N > p$ でなければならない．

ここで $g(x)$ は，リンク関数（link function）と呼ばれる 1 対 1 の単調な変換で

ある．実際の解析では，リンク関数はデータへの当てはまりに応じて選択される．指数型分布族の性質から明らかなように，θ_i，$\mathrm{E}(Y_i|\boldsymbol{X}_i)$，$\boldsymbol{\beta}$ は，関数 $b(x)$，$c(x)$，$g(x)$ を用いて，相互に変換可能な関係にある．

■ 10.2.2 十分統計量と正準リンク関数

一般化線型モデルの同時分布は積の形式になっているが，これは最尤法の計算上都合がよい．なぜなら対数尤度関数が

$$l(\theta_1,...,\theta_N) = \sum_{i=1}^{N} y_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)$$

という関数 $b(x)$，$c(x)$，$d(x)$ の和で表されるからである．

リンク関数として，データの特徴や目的に応じてさまざまな変換を用いることができるが，特に重要なのは，対数尤度の第 1 項が

$$\sum_{i=1}^{N} y_i b(\theta_i) = \sum_{i=1}^{N} y_i \boldsymbol{X}_i \boldsymbol{\beta}$$

と表されるような関数である．なぜなら，このときパラメータ $\boldsymbol{\beta}$ の最尤推定量が，$\sum_{i=1}^{N} y_i \boldsymbol{X}_i$ だけに依存することになるからである．このとき，特にサンプルサイズが小さいときに計算が安定することが知られている．$\sum_{i=1}^{N} y_i \boldsymbol{X}_i$ を一般化線型モデルの十分統計量（sufficient statistics）という[*1)]．そして，上のような性質を満たすリンク関数は正準リンク関数（canonical link function）と呼ばれ，実際の統計解析ではこれを選ぶことが多い．

$$\mathrm{E}(Y_i|\boldsymbol{X}_i) = \mu_i$$

とおけば，正規分布，2 項分布，Poisson 分布の正準リンク関数は

$$g(\mu_i) = \mu_i \qquad \text{for normal}$$

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) \qquad \text{for binomial}$$

$$g(\mu_i) = \log(\mu_i) \qquad \text{for Poisson}$$

となる．正準リンク関数を用いたモデルは，他のリンク関数に比べて，サンプ

[*1)] ある統計量 $t(\boldsymbol{Y})$ が，パラメータ θ の十分統計量であるとは，$t(\boldsymbol{Y})$ を与えたときの $\boldsymbol{Y}$ の条件付分布が，θ の値に依存しなくなることをいう．
十分統計量とは，ある意味でデータに含まれる情報をすべて要約するような統計量のことである．一般化線型モデルのケースでは，データの持つパラメータ $\boldsymbol{\beta}$ に関する情報は，$\sum_{i=1}^{N} y_i \boldsymbol{X}_i$ にすべて含まれている．

ルサイズが小さいとき最尤推定量の挙動がよい．そのよい例が，2 項分布におけるロジスティック回帰である．

これまでの結果を整理しよう．表 10-1 のように正規分布，2 項分布，Poisson 分布はすべて指数型分布族の形に変形できるから，一般化線型モデルに拡張することができる．そしてそれぞれが対応する正準リンク関数を持つ．

表 10-1　指数型分布族の例

確率分布	$b(\theta)$	$c(\theta)$	$d(y)$	正準リンク関数
正規分布	$\frac{\mu}{\sigma^2}$	$-\frac{\mu^2}{2\sigma^2}$	$-\frac{y^2}{2\sigma^2}-\frac{1}{2}\log(2\pi\sigma^2)$	$g(x)=x$
2 項分布	$\log\left(\frac{\pi}{1-\pi}\right)$	$N\log(1-\pi)$	$\log\binom{N}{y}$	$g(x)=\log\left(\frac{x}{1-x}\right)$
Poisson 分布	$\log(\lambda)$	$-\lambda$	$-\log(y!)$	$g(x)=\log(x)$

■ 10.2.3　推　定

一般化線型モデルにおいて，パラメータベクトル $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)^T$ の最尤推定量はどのように計算されるのだろうか．その方針はこれまでと同様である．つまり，対数尤度を偏微分した β_j のスコア関数

$$U_j=\frac{\partial l(\theta_1,\ldots,\theta_N)}{\partial\beta_j}$$

の具体的な形を求めて，パラメータ $\boldsymbol{\beta}$ 全体のスコア方程式

$$\boldsymbol{U}(\boldsymbol{\beta})=\begin{pmatrix}U_1\\ \vdots\\ U_p\end{pmatrix}=0$$

を解けばよい．

導出は後に述べることにして，先に結果を示そう．パラメータ β_j に対応するスコア関数の要素は

$$U_j=\sum_{i=1}^{N}\frac{y_i-\mu_i}{\mathrm{Var}(Y_i|\boldsymbol{X}_i)}x_{ij}\frac{\partial\mu_i}{\partial\eta_i}$$

で与えられる．ただし

$$\mathrm{E}(Y_i|\boldsymbol{X}_i)=\mu_i$$

$$g(\mu_i)=\eta_i$$

とおいた.

Fisher 情報行列 $\boldsymbol{I} = \mathrm{E}(UU^T)$ は，$I_{jk} = \mathrm{E}(U_j U_k)$ を j 行目 k 列目の要素とする $N \times N$ 行列になる．この要素は，スコア関数を微分することで

$$I_{jk} = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{\mathrm{Var}(Y_i|\boldsymbol{X}_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

と導ける．全体を行列として表すときは，$\boldsymbol{W}$ を

$$w_{ii} = \frac{1}{\mathrm{Var}(Y_i|\boldsymbol{X}_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

を i 行 i 列の要素とする $N \times N$ 対角行列として定義するよい．Fisher 情報行列は

$$\boldsymbol{I} = \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}$$

と表すことができる.

最後に，パラメータ β_j のスコア関数がどのように導かれたのかを示してこの節を終わろう．一般化線型モデルにおいて，対数尤度関数は個々のデータの和の形になる．そのため，i 番目のデータがどのように尤度へ貢献するかは

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

と表される．β_j のスコア関数は，微分の連鎖公式を用いて

$$U_j = \sum_{i=1}^{N} \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{N} \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j}$$

と表すことができる．3 つの偏微分についてひとつひとつ考えていこう．目標は $b(\theta_i)$，$c(\theta_i)$，$d(y)$ の部分を具体的な形に置き換えることである．ひとつ目の部分は

$$\frac{\partial l_i}{\partial \theta_i} = y_i b'(\theta_i) + c'(\theta_i) = b'(\theta_i)(y_i - \mu_i)$$

となる．次に

$$\mu_i = -\frac{c'(\theta_i)}{b'(\theta_i)}$$

を微分して

$$\frac{\partial \mu_i}{\partial \theta_i} = \frac{-c''(\theta_i)}{b'(\theta_i)} + \frac{c'(\theta_i) b''(\theta_i)}{b'(\theta_i)^2} = b'(\theta_i)\mathrm{Var}(Y_i|\boldsymbol{X}_i)$$

となることを利用すれば

$$\frac{\partial \theta_i}{\partial \mu_i} = \left(\frac{\partial \mu_i}{\partial \theta_i} \right)^{-1} = \frac{1}{b'(\theta_i)\mathrm{Var}(Y_i|\boldsymbol{X}_i)}$$

が導かれる．最後の部分は，$\eta_i = \boldsymbol{X}_i \boldsymbol{\beta}$ の偏微分を考えれば

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij}$$

となる．個々のデータのスコアへの貢献は，3 つの偏微分の積である．よって総和をとれば

$$U_j = \sum_{i=1}^{N} \frac{y_i - \mu_i}{\mathrm{Var}(Y_i|\boldsymbol{X}_i)} x_{ij} \frac{\partial \mu_i}{\partial \eta_i}$$

が導かれる．

Newton–Raphson 法

スコア方程式を解くにはどうすればよいか．一般化線型モデルのスコア関数は，$\boldsymbol{\beta}$ の非線型の関数になることがふつうなので，簡単に解を求めることができない．そこで，Newton–Raphson（ニュートン・ラプソン）法などの計算アルゴリズムを用いることが一般的である．

Newton–Raphson 法は，ある関数 $f(x)$ とその導関数 $f'(x)$，計算の初期値 $x^{(a=0)}$ が与えられたとき，方程式 $f(x) = 0$ の解を求めるアルゴリズムである．以下の公式を用いた反復計算によって x を更新する．

$$x^{(a+1)} = x^{(a)} - \frac{f\left(x^{(a-1)}\right)}{f'\left(x^{(a-1)}\right)}$$

ここで a は反復回数であり，左辺は a 番目の計算の解を表している．

Newton–Raphson 法では，反復計算を止めるための基準が必要になる．たとえば正の実数 $\varepsilon > 0$ を指定して，反復計算ごとに関数がじゅうぶんゼロに近く

$$|f(x^{(a)})| < \varepsilon$$

となるかをみる，あるいは x の変化が小さく

$$|x^{(a)} - x^{(a-1)}| < \varepsilon$$

を満たすか判定する，といった手続きがとられる．

Newton–Raphson 法をコンピューターで実行するとき，計算精度は浮動小数点の桁数で決まっているから，計算ごとに丸め誤差が生じる．そのため，方程式 $f(x) = 0$ を解くことは，$f(x)$ が十分ゼロに近い x を探すことと考えて実用上は差し支えない．

10.3 情報量規準とモデルの選択

■ 10.3.1 赤池情報量規準

尤度比検定は，対数尤度が帰無仮説と対立仮説のどちらのモデルを支持しているかを判定する手法である．これに対して，解析に用いたモデルは間違っているかもしれないが，どれくらいデータに適合しているかを評価したいときがある．たとえば，一般化線型モデルによる解析では，デザイン行列の異なるモデルの候補は無数にある．このとき，どのモデルを解析に用いるべきかデータへの当てはまりに基づいて選択しなければならない．

モデル選択の指標として，対数尤度関数に最尤推定量を代入した $l(\widehat{\boldsymbol{\beta}})$ を用いることは好ましくない．なぜなら，$l(\widehat{\boldsymbol{\beta}})$ には複雑なモデルを選ぶ方向にバイアスがあるからである．具体例として，一部のパラメータがゼロになるような 2 つの一般化線型モデルで説明しよう．

$$\text{Model A}: \quad \boldsymbol{\beta} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\beta}_1 \end{pmatrix}$$

$$\text{Model B}: \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{pmatrix} \neq \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\beta}_1 \end{pmatrix}$$

このとき，仮にモデル A が正しかったとしても（つまり $\boldsymbol{\beta}_0 = 0$ だとしても），複雑なモデル B を当てはめたら，得られる推定量 $\widehat{\boldsymbol{\beta}}_0$ はゼロにはならず，対数尤度がもっとも大きくなるような値が選ばれる．これは，対数尤度の値を大きくするような偏りを生じさせる．これを過適合（overfitting）といって，最尤法の欠点のひとつとされている．実際，モデル A とモデル B の対数尤度の値を比べると，モデル B の方が常に大きくなる．つまり，モデル選択の指標として対数尤度関数に最尤推定量を代入した値を用いると，増えたパラメータに意味がなかったとしても，パラメータの数が大きいモデルを選んでしまう[*2)]．

モデル選択の指標は数多くあるが，ここでは対数尤度から簡便に計算できる赤池情報量規準（AIC）を紹介する（Akaike 1973）．この指標は，最尤推定量の下で評価した対数尤度関数を用いて

[*2)] モデル選択のため 9 章の決定係数（R^2）を参考にすることもあるが，過適合の問題を伴うのは，対数尤度と同じである．

$$\text{AIC for Model A} = -2l(\widehat{\boldsymbol{\beta}}_{\text{A}}) + 2q$$

や

$$\text{AIC for Model B} = -2l(\widehat{\boldsymbol{\beta}}_{\text{B}}) + 2p$$

と定義される．ただし $\widehat{\boldsymbol{\beta}}_{\text{A}} = (\mathbf{0}, \widehat{\boldsymbol{\beta}}_1)^T$ と $\widehat{\boldsymbol{\beta}}_{\text{B}}$ はそれぞれモデル A と B の最尤推定量，q と p はそれぞれのパラメータ数である．モデルの候補が複数あるとき，AIC が小さいほど予測性能がよいモデルと判断される．

■ 10.3.2　赤池情報量規準と Kullback–Leibler 情報量の関係

データに当てはめたモデル $p(y;\boldsymbol{\beta})$ と真のモデル $q(y)$ との距離を測る指標のひとつとして，Kullback–Leibler 情報量

$$KL = \int \log\left[\frac{q\,(y)}{p\,(y;\boldsymbol{\beta})}\right] q\,(y)\,dy$$

がある．モデル選択では AIC が小さいモデルがよいと判断されるわけだが，それは Kullback–Leibler 情報量が小さいという意味で，真のモデルに近いモデルを選択するという操作になっている．この点について補足しよう．

Kullback–Leibler 情報量は

$$KL = \int \log\left[q\,(y)\right] q\,(y)\,dy - KL(\widehat{\boldsymbol{\beta}})$$

というように当てはめたモデル $p(y;\widehat{\boldsymbol{\beta}})$ に依存しない項と依存する項にわけることができる．ただし

$$KL\,(\boldsymbol{\beta}) = \text{E}[\log[p(y;\boldsymbol{\beta})]]$$

とおいた．第 1 項はモデルに依存しないから，$\text{KL}(\widehat{\boldsymbol{\beta}})$ が最大になるようなモデルが，Kullback–Leibler 情報量を最小にするわけである．ただし対数尤度は $l(\boldsymbol{\beta}) = \sum_{i=1}^{N} \log\left[p(y_i;\boldsymbol{\beta})\right]$ と定義されるから，対数尤度と $KL(\widehat{\boldsymbol{\beta}})$ には

$$KL\,(\boldsymbol{\beta}) = \frac{1}{N}\text{E}[l(\boldsymbol{\beta})]$$

という関係があることに注意してほしい．

$KL(\widehat{\boldsymbol{\beta}})$ を，Taylor 展開などを用いて近似すると，以下の結果が得られる．AIC との関係をわかりやすくするため，-2 を掛けたものを考えると

$$\begin{aligned}
-2N \times KL(\widehat{\boldsymbol{\beta}}) &\approx -2N \times KL\,(\boldsymbol{\beta}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \boldsymbol{I}\,(\boldsymbol{\beta})\,(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&\approx -2\text{E}[l(\boldsymbol{\beta})] + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \boldsymbol{I}\,(\boldsymbol{\beta})\,(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&\approx -2\text{E}[l(\widehat{\boldsymbol{\beta}})] + 2(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \boldsymbol{I}\,(\boldsymbol{\beta})\,(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&\approx -2\text{E}[l(\widehat{\boldsymbol{\beta}})] + 2\chi_p^2
\end{aligned}$$

さらに両辺の期待値をとると，$\mathrm{E}(\chi^2_p) = p$ だから以下の結果が得られる．

$$-2N \times \mathrm{E}[KL(\widehat{\boldsymbol{\beta}})] \approx -2\mathrm{E}[l(\widehat{\boldsymbol{\beta}})] + 2p$$

次に AIC の期待値を考えると，これは

$$\mathrm{E}\,(AIC) = -2\mathrm{E}[l(\widehat{\boldsymbol{\beta}})] + 2p$$

であることは定義から明らかである．両者の式を比べてみると，$-2N \times KL(\widehat{\boldsymbol{\beta}})$ と AIC の期待値には近似的な関係があることがわかる．この結果は，あるモデルの AIC が小さいとき，Kullback–Leibler 情報量の意味で真のモデルに近いであろうことを意味している．

この議論は期待値のみ考えていて，AIC のバラツキを考えていないことに注意してほしい．あるモデルが別のモデルに比べて，AIC がわずかに小さいだけでは，Kullback–Leibler 情報量に真に差はないこともあり得る．このようなケースでは，モデル間に予測性能の差はないと判断すべきだろう．

■ 10.3.3 事例：大気汚染物質と死亡率

AIC を用いたモデル選択について，6 都市研究データを例に説明しよう．ここでは，8 種類の大気汚染物質のうちどれがもっとも死亡率との関連が強いか，そしてその関連は直線でじゅうぶん説明できるか（2 次以上の項に意味があるか）に関心があるとする．

データをみて，まず気づくことは大気汚染物質濃度間に強い相関があるということである．図 10-1 は，微小粒子と硫酸塩粒子の散布図である．両者の相関係数は 0.98 と非常に強い．微小粒子と硫酸塩粒子を，同時に共変量に含めたときの分散拡大因子[*3)] は 33.3 と非常に大きい．さらに，SO_2 濃度は，総粒子，微小粒子，吸入性粒子，エアロゾル酸度の数値の線型結合によって計算できる．したがって，これらの大気汚染物質を複数デザイン行列に含めたとしても，多重共線性[*3)] が生じたり，回帰係数が一意に定まらなかったりするから，有用な情報は得られないだろう．

そこで，6 都市研究の 8 種類の大気汚染物質・死亡率データに，以下のような Poisson 回帰モデルを当てはめ，AIC を用いてモデルを比較する．モデルの候補は以下の 16 通りになる．

[*3)] 多重共線性や分散拡大因子については 9.9 節参照.

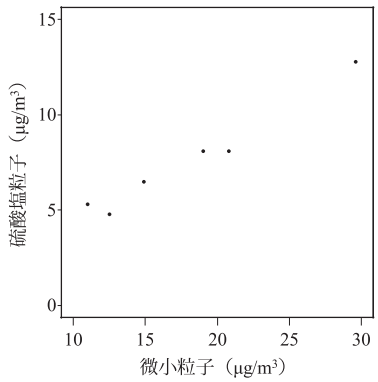

図 10-1　6 都市研究における微小粒子と硫酸塩粒子の散布図

$$\log[\mathrm{E}(Y|\text{TOTAL PARTICLE})] = \text{INTERCEPT} + \text{TOTAL PARTICLE}$$

$$\log[\mathrm{E}(Y|\text{TOTAL PARTICLE})] = \text{INTERCEPT} + \text{TOTAL PARTICLE} + \text{TOTAL PARTICLE}^2$$

$$\log[\mathrm{E}(Y|\text{INHALABLE PARTICLE})] = \text{INTERCEPT} + \text{INHALABLE PARTICLE}$$

$$\log[\mathrm{E}(Y|\text{INHALABLE PARTICLE})] = \text{INTERCEPT} + \text{INHALABLE PARTICLE} + \text{INHALABLE PARTICLE}^2$$

$$\log[\mathrm{E}(Y|\text{FINE PARTICLE})] = \text{INTERCEPT} + \text{FINE PARTICLE}$$

$$\log[\mathrm{E}(Y|\text{FINE PARTICLE})] = \text{INTERCEPT} + \text{FINE PARTICLE} + \text{FINE PARTICLE}^2$$

$$\log[\mathrm{E}(Y|\text{FINE PARTICLE})] = \text{INTERCEPT} + \text{SULFATE PARTICLE}$$

$$\log[\mathrm{E}(Y|\text{SULFATE PARTICLE})] = \text{INTERCEPT} + \text{SULFATE PARTICLE} + \text{SULFATE PARTICLE}^2$$

$$\log[\mathrm{E}(Y|\text{ARE ROSOL ACIDITY})] = \text{INTERCEPT} + \text{ARE ROSOL ACIDITY}$$

$$\log[\mathrm{E}(Y|\text{ARE ROSOL ACIDITY})] = \text{INTERCEPT} + \text{ARE ROSOL ACIDITY} + \text{ARE ROSOL ACIDITY}^2$$

$$\log[\mathrm{E}(Y|\text{SULFUR DIOXIDE})] = \text{INTERCEPT} + \text{SULFUR DIOXIDE}$$

$$\log[\mathrm{E}(Y|\text{SULFUR DIOXIDE})] = \text{INTERCEPT} + \text{SULFUR DIOXIDE} + \text{SULFUR DIOXIDE}^2$$

$$\log[\mathrm{E}(Y|\text{NIROGEN DIOXIDE})] = \text{INTERCEPT} + \text{NIROGEN DIOXIDE}$$

$$\log[\mathrm{E}(Y|\text{NIROGEN DIOXIDE})] = \text{INTERCEPT} + \text{NIROGEN DIOXIDE} + \text{NIROGEN DIOXIDE}^2$$

$$\log[\mathrm{E}(Y|\text{OZONE})] = \text{INTERCEPT} + \text{OZONE}$$

$$\log[\mathrm{E}(Y|\text{OZONE})] = \text{INTERCEPT} + \text{OZONE} + \text{OZONE}^2$$

表 10-2 は，16 通りの Poisson 回帰モデルを当てはめ，AIC を求めた結果である．AIC が小さい（適合度がよい）のは SO_2（2 次曲線），NO_2（2 次曲線），SO_2（直線）の順である．この 3 つのモデルの AIC の差は小さく，データへの当てはまりは同程度である．総粒子と吸入性粒子は，直線の方が，AIC は小さい．これは，2 次の項を追加しても，適合度が改善しなかったことを意味している．

表 10-3 と図 10-2 に，もっとも AIC が小さかった SO_2 濃度に関する推定結果と予測曲線を示す．図のドットは各都市の死亡率の実測値であり，実線は直線モデル，破線は 2 次曲線モデルである．6 都市研究では，大気汚染物質は都市単位で測定された．そのため，図 10-2 の実測値（ドット）は 6 点しかない．都市間の死亡率の違いは，横軸に SO_2 濃度をとった 2 次曲線でよく説明されている．

表 10-2 6 都市研究データにおける AIC による直線・2 次曲線モデルの比較

モデル	AIC	モデル	AIC
SO_2（2 次曲線）	53.5	微小粒子（2 次曲線）	63.8
NO_2（2 次曲線）	56.0	微小粒子（直線）	64.3
SO_2（直線）	56.8	総粒子（2 次曲線）	65.4
硫酸塩粒子（2 次曲線）	57.7	吸入性粒子（直線）	68.2
NO_2（直線）	58.5	吸入性粒子（2 次曲線）	70.0
オゾン（2 次曲線）	59.5	オゾン（直線）	75.9
硫酸塩粒子（直線）	61.9	エアロゾル濃度（2 次曲線）	91.7
総粒子（直線）	63.5	エアロゾル濃度（直線）	94.3

表 10-3 6 都市研究データにおける SO_2 濃度と死亡率の関係を表す直線・2 次曲線モデルの比較

	直線				2 次曲線			
	回帰係数	95%信頼区間		p 値	回帰係数	95%信頼区間		p 値
切片	−4.573	−4.661	−4.485	< 0.01	−4.712	−4.862	−4.562	< 0.01
SO_2（1 次）	0.214	0.149	0.280	< 0.01	0.549	0.256	0.843	< 0.01
SO_2（2 次）					−0.013	−0.023	−0.002	0.02

*回帰係数は SO_2 10 ppb（またはその 2 乗）増加あたりの値を示す．

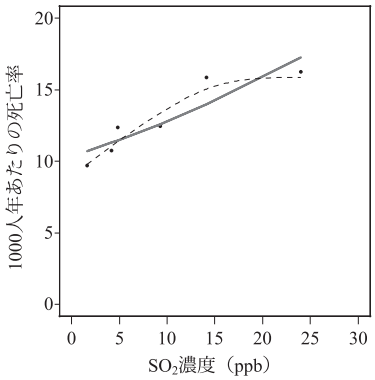

図 10-2　6 都市研究データにおける SO_2 濃度と死亡率の関係を表す散布図
ドットは各都市の死亡率の実測値，実線は直線 Poisson 回帰，破線は 2 次曲線 Poisson 回帰

■ 10.3.4　事例から得られた教訓

表 10-2 の結果は，2 次曲線が，大気汚染物質と死亡率の関係を完全に表す真のモデルということを意味しているわけではない．なぜなら AIC は，モデル間の相対的な評価にすぎないからである．モデルがどんなにデータに当てはまっていたとしても，年齢，性，疾患といった死亡に関連することがわかっている個人レベルの共変量は考慮されていない．このように考えると，解析に用いられたデータは都市レベルのもので，死亡という個人レベルの現象をモデル化するには限界があることがわかるだろう．

6 都市研究では，個人レベルの共変量（年齢，性，喫煙，教育歴，BMI，職業上の曝露，高血圧，糖尿病）を調整してもなお，微小粒子と死亡率に関連があったと報告されている．論文の結論は，「未測定で，検討されていないリスク因子の影響を排除できたとはいえないが，微小粒子や他の複合的な汚染物質は米国のいくつかの都市の死亡率の増加に寄与していることを，この結果は示唆している」というものだった（Dockery, et al. 1993）．

公害問題からの教訓

6 都市研究は，大気汚染といういわゆる公害問題を扱った事例である．疫学研究を行って，公害による健康被害の有無やその程度・範囲を調べることは多い．それは医学研究の中で，もっとも困難が多くデリケートなもののひとつである．実際に多くの公害問題にかかわった統計学者の発言がある（吉村 1976）．

「公害においてまず問題になることは，異常の確認と原因の追及である．公害は人為的災害である．すなわち，加害者と被害者があり，前者はきわめて横着である．被害者が直接の被害感にもとづき，直観で感じられる加害者に発生源除去を申し入れても，全く相手にしないのがふつうである．

（中略）いろんな分野の熟達者が，専門にとらわれずに対象となっている問題の解決に協力し，努力するのが役に立つ．その際，統計的視点はかけてはならないものである．（中略）視点として特に必要なものは

- コントロールを適切に設定すること
- 測定の誤差と対象そのものの持つ変動性とを区別し，かつその大きさを評価すること
- 適切な分布法則を想定すること
- 偏りとばらつきを区別し，前者によって推論がゆがめられないようにすること
- 層別などによってみかけの相関を除くこと
- 変数変換などを駆使して，単純な関数関係の表現を見出すこと
- 誤差的変動をこえて出される法則性，つまり，統計的有意ということの実質的意味を誤解しないこと
- 統計的に認識されるうわべの関係や特徴を，内在する法則性への確信に転化させる論理を正しく用いること

などである」

これは 50 年近く前の発言だが，最近の医学研究でもその重要性は変わっておらず，普遍性のある教訓と思う．本書を執筆するにあたって，それぞれの視点をできるだけ盛り込むように心掛けた．

演 習 問 題

〈指数型分布族の期待値〉

問 1　分散が既知の正規分布に従う確率変数 Y の期待値を求めたい．この場合の $b'(\mu), c'(\mu)$，$-c'(\mu)/b'(\mu)$ の組み合わせとして，正しいものを選べ．ただし，$b'(\mu)$ は $b(\mu)$ の，$c'(\mu)$ は $c(\mu)$ の導関数である．

(A) $b'(\mu) = 1$　　$c'(\mu) = -\mu$　　$-c'(\mu)/b'(\mu) = \mu$
(B) $b'(\mu) = \mu$　　$c'(\mu) = -\mu^2$　　$-c'(\mu)/b'(\mu) = \mu$
(C) $b'(\mu) = 1/\sigma^2$　　$c'(\mu) = -\mu/\sigma^2$　　$-c'(\mu)/b'(\mu) = \mu$
(D) $b'(\mu) = \mu/\sigma^2$　　$c'(\mu) = -\mu^2/\sigma^2$　　$-c'(\mu)/b'(\mu) = \mu$

問 2　Poisson 分布に従う確率変数 Y の期待値を求めたい．この場合の $b'(\mu)$，$c'(\mu)$，$-c'(\mu)/b'(\mu)$ の組み合わせとして，正しいものを選べ．ただし

$$b(\lambda) = \log(\lambda T)$$

$$c(\lambda) = -\lambda T$$

であり，$b'(\lambda)$ は $b(\lambda)$ の，$c'(\lambda)$ は $c(\lambda)$ の導関数である．

(A) $b'(\lambda) = 1/\lambda$　$c'(\lambda) = -T$　$-c'(\lambda)/b'(\lambda) = \lambda T$
(B) $b'(\lambda) = 1/T$　$c'(\lambda) = -T$　$-c'(\lambda)/b'(\lambda) = T^2$
(C) $b'(\lambda) = \lambda T$　$c'(\lambda) = -T$　$-c'(\lambda)/b'(\lambda) = 1/\lambda$
(D) $b'(\lambda) = T/\lambda$　$c'(\lambda) = -T$　$-c'(\lambda)/b'(\lambda) = \lambda$

11 正規線型モデル

この章では，分散分析・回帰分析の現代版である正規線型モデルについて述べる．古典的な回帰係数の推定方法である最小 2 乗推定量は，最尤推定量として導出することができる．降圧薬臨床試験データを例に，ベースライン値のあるランダム化臨床試験におけるコントロール群の必要性や共変量調整の意義について述べる．糸球体濾過率とクレアチニン濃度の関係を例に，残差プロットが，非線型によるモデルの誤特定を防ぐために有用であることを示す．

キーワード 一般線型モデル，共変量調整，正規線型モデル，正規分布，最小 2 乗法，残差プロット，線型性，平均への回帰

事　例 降圧薬臨床試験，糸球体濾過率研究

11.1 モデルの構造

アウトカム Y_i（$i = 1,\ldots,N$）が，平均が異なる独立な正規分布に従い，平均と共変量 $X_1, X_2, \ldots, X_p$ の関係が，恒等リンク関数を介して線型の関係にあるとき，これを一般線型モデル（general linear model）または正規線型モデルという．

$$Y_i \sim N(\mu_i, \sigma^2)$$

$$\mathrm{E}(Y_i|\boldsymbol{X}_i) = \mu_i = \boldsymbol{X}_i\boldsymbol{\beta}$$

もちろんこれは一般化線型モデルの一種であり，分散分析・回帰分析もこのモデルに含まれる．

11.2　推　　定

分散が既知のときの対数尤度関数は，正規分布の密度関数から

$$l(\boldsymbol{\beta},\sigma) = \sum_{i=1}^{N} \frac{-1}{2} \left(\frac{y_i - \boldsymbol{X}_i\boldsymbol{\beta}}{\sigma} \right)^2$$

となる．これは，Y_i と $\boldsymbol{X}_i\boldsymbol{\beta}$ の差の 2 乗だから，対数尤度を最大化する操作は，9 章の最小 2 乗解を求める計算と同じものである．すでに述べたように，対数尤度を最大にする値は，（逆行列が存在すればそれを用いて）

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

という明示的な解を導くことができる．一方で，分散の推定には，不偏推定量

$$\widehat{\sigma}^2 = \frac{1}{N-p-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$$

を用いることが一般的である．

正規線型モデルにおいて，最尤推定量 $\widehat{\boldsymbol{\beta}}$ は

$$\widehat{\boldsymbol{\beta}} \overset{d}{\to} N[\boldsymbol{\theta},\sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}]$$

という漸近正規性を持つだけではなく，たとえ小標本でも不偏性

$$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

が成り立っている．

11.3　事例：非線型性によるモデルの誤特定

正規線型モデルは，糸球体濾過率や血圧といった臨床検査で得られた測定値の統計解析のためによく用いられる．Brochner–Mortensen et al.（1977）は，男性 180 人と女性 200 人を対象に，血漿クレアチニン濃度と糸球体濾過率の関係を調べた．表 11-1 は，ランダムに選ばれた対象者 31 人のデータであり，図 11-1 左上のグラフはこのデータから描いた散布図である．血漿クレアチニン濃度を X_i，糸球体濾過率を Y_i とすると，両者には単調な関係がみられている．

最小 2 乗法により一次関数のモデル

$$\mathrm{E}(Y_i|X_i) = \beta_0 + \beta_1 X_i$$

表 11-1 糸球体濾過率研究データ

対象	糸球体濾過率 (mL/min)	クレアチニン (mg/dL)	対象	糸球体濾過率 (mL/min)	クレアチニン (mg/dL)
1	90	0.85	17	38	1.83
2	45	0.99	18	47	1.98
3	103	1.13	19	45	2.03
4	100	1.13	20	40	2.09
5	93	1.13	21	27	2.77
6	90	1.13	22	37	2.96
7	70	1.13	23	25	3.11
8	77	1.27	24	15	3.96
9	47	1.41	25	15	4.69
10	45	1.47	26	20	4.8
11	60	1.47	27	10	5.93
12	53	1.56	28	5	5.93
13	35	1.69	29	5	5.93
14	63	1.7	30	10	7.79
15	55	1.75	31	12	11.02
16	35	1.75			

を当てはめると，$\widehat{\beta_0} = 71.1$，$\widehat{\beta_1} = -9.0$ となる．図 11-1 右上は散布図にこの直線を加えたものである．この一次関数モデルは，データによく当てはまっているといってよいだろうか．

図 11-1 左下は，横軸に X_i を，縦軸に残差

$$Y_i - \widehat{\mathrm{E}}(Y_i|X_i) = Y_i - 71.1 + 9.0X_i$$

をとった残差プロットである．よくみると，残差とクレアチニン濃度の関係に U 字型の傾向が見つかる．これは線型性（linearity）の仮定が間違っていることを意味している．

生理学的に考えると，両者の関係は直線ではない．反比例の関係，つまり糸球体濾過率が半分になると，クレアチニン濃度は 2 倍になるような対応があるはずである．これは

$$\mathrm{E}(Y_i|X_i) = \beta_0 + \frac{\beta_1}{X_i}$$

という逆数のモデルが正しいことを意味する．このモデルを当てはめるのは難しくない．クレアチニン濃度の逆数を計算して，それを説明変数とした最小 2 乗法を行えばよい．その結果，回帰係数は $\widehat{\beta_0} = -2.5$，$\widehat{\beta_1} = 88.3$ となる．この逆数のモデルを，もともとの散布図に図示すると図 11-1 右下のようになる．このグラフからは，クレアチニン濃度を逆数に変数変換することで，データへの

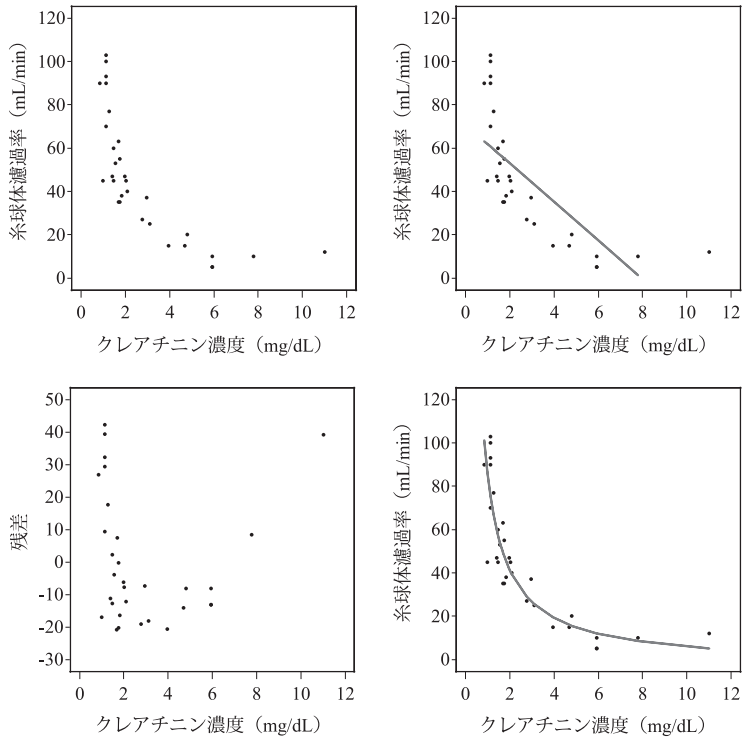

図 11-1　糸球体濾過率データにおける糸球体濾過率と血漿クレアチニン濃度の散布図（左上，右上，右下）と一次関数モデルを当てはめたときの残差プロット（左下）

当てはまりが改善していることがわかる．

一次関数モデルと逆数モデルのデータへの当てはまりを比べてみよう．それぞれのモデルの AIC は 278.2 と 252.8 である．AIC でみても逆数モデルを選択すべきことがわかる[*1)]．

11.4　事例：ベースライン値のある臨床試験の解析 2

次に述べるのは，臨床試験でベースライン値をどのように扱うかという事例である．表 11-2 は，アルブミン尿を呈する 1 型糖尿病・高血圧患者 16 人に，カ

[*1)] 検査方法は当時と異なるが，現代の医療でも，糸球体濾過率を測るためにクレアチニンからの推定値が利用されている．その計算でも逆数に近い数式が用いられている．

表 11-2 降圧薬臨床試験データ

治療	ベースライン収縮期血圧（mmHg）	1 週目の収縮期血圧（mmHg）	治療	ベースライン収縮期血圧（mmHg）	1 週目の収縮期血圧（mmHg）
カプトプリル	147	137	プラセボ	133	139
カプトプリル	129	120	プラセボ	129	134
カプトプリル	158	141	プラセボ	152	136
カプトプリル	164	137	プラセボ	161	151
カプトプリル	134	140	プラセボ	154	147
カプトプリル	155	144	プラセボ	141	137
カプトプリル	151	134	プラセボ	156	149
カプトプリル	141	123			
カプトプリル	153	142			

プトプリルまたはプラセボをランダムに割付けた臨床試験である（Hommel, et al. 1986）．この論文では，カプトプリル群 9 人では治療開始して 1 週目の収縮期血圧は，ベースライン値に比べて有意に低下し（対応のある t 検定 $p < 0.01$），プラセボ群 7 人では有意な変化はなかったため（$p = 0.17$），カプトプリルは有効と述べられている．この結論は正しいだろうか．

このようなベースライン値のあるランダム化臨床試験では，治療前後の変化量をアウトカムにすることができる．しかしベースラインから有意な変化があったかどうかを主たる解析にすべきではない．治療前に重症な患者は，ランダムな変動や自然軽快によって改善する現象，すなわち平均への回帰（regression to mean）が生じるからである．治療が有効かどうかの判断は，変化の有無ではなく，治療間でアウトカムに差があったかどうかによってなされなければならない．

治療前後の変化量をアウトカムとした正規線型モデルは

$$\text{E(CHANGE in SBP|TREATMENT)} = \text{INTERCEPT} + \text{TREATMENT}$$

と表すことができる．このモデルに対応するデータを行列表現すると

$$\boldsymbol{Y} = \begin{pmatrix} 137 - 147 \\ \vdots \\ 142 - 153 \\ 139 - 133 \\ \vdots \\ 149 - 156 \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ \vdots & 1 \\ \vdots & 0 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix}$$

となる．変化量のモデルが，ベースライン値を用いない解析

$$\text{E(SBP at 1 WEEK|TREATMENT)} = \text{INTERCEPT} + \text{TREATMENT}$$

より優れているのは，どのようなときだろうか．この場合に対応するデータは以下のようなものである．

$$\boldsymbol{Y} = \begin{pmatrix} 137 \\ \vdots \\ 142 \\ 139 \\ \vdots \\ 149 \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ \vdots & 1 \\ \vdots & 0 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix}$$

変化量のモデルの方がよい結果をもたらすための条件は，以下のように調べることができる．収縮期血圧のベースライン値，治療後の測定値，変化量の分散をそれぞれ，σ_0^2，σ_1^2，σ_2^2 で表す．ベースライン値と治療後の測定値の相関を ρ とする．変化量の分散は（いわゆる確率変数の和の分散だから）

$$\sigma_2^2 = \sigma_0^2 + \sigma_1^2 - 2\rho\sigma_0\sigma_1 = \sigma_1^2\left(\frac{\sigma_0^2}{\sigma_1^2} + 1 - \frac{2\rho\sigma_0}{\sigma_1}\right)$$

と表すことができる．変化量のモデルを用いる目的は，個人によってベースライン値は異なっているから，それを引くことで個人間のバラツキを減らすためである．そこで σ_2^2 が σ_1^2 より小さくなる条件を調べると

$$\sigma_1^2\left(\frac{\sigma_0^2}{\sigma_1^2} + 1 - \frac{2\rho\sigma_0}{\sigma_1}\right) < \sigma_1^2 \qquad \text{if } \rho > \frac{\sigma_0/\sigma_1}{2}$$

が得られる．これは，分散が $\sigma_0^2 = \sigma_1^2$ というように等しいとき，治療前後の相関が 0.5 より高ければ，変化量をとることで分散を減らせることを意味している．

変化量をアウトカムにするのではなく，ベースライン値を共変量とすることもできる．このモデルは共分散分析（analysis of covariance）と呼ばれており

$$\text{E(SBP at 1 WEEK|TREATMENT, SBP at BASELINE)}$$
$$= \text{INTERCEPT} + \text{TREATMENT} + \text{SBP at BASELINE}$$

$$\boldsymbol{Y} = \begin{pmatrix} 137 \\ \vdots \\ 142 \\ 139 \\ \vdots \\ 149 \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & 1 & 147 \\ \vdots & \vdots & \vdots \\ \vdots & 1 & 153 \\ \vdots & 0 & 133 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 156 \end{pmatrix}$$

と表される．臨床試験の文脈では，ベースライン値を共変量としてモデルに含める解析のことを，共変量調整（covariate adjustment）と呼んでいる．

以上の結果をまとめよう．要点は 2 つある．ベースライン値のある臨床試験では，治療前後の変化量をアウトカムにすることができるが，平均への回帰を避けるため，有意な変化があったかどうかではなく，ランダム化した比較対照をおくべきである．そして，治療前後の測定値間の相関が高ければ，変化量の解析や共変量調整を採用するべきである．ここでいう共変量調整は，推定精度や検出力の向上のための手段であって，バイアスの排除が目的ではない．

表 11-3 に上で述べた 3 つのモデルで解析した結果を示す．このデータでは相関係数は 0.60 であり，ベースライン値によって検出力の上昇が期待できる．ただし，この場合は平均の比較と変化量の解析で，95%信頼区間の幅に大きな違いはない．カプトプリル群とプラセボ群の有意な差がみられたのは，3 つの手法のうちもっとも検出力が高い共分散分析だけだった．

表 11-3 降圧薬試験データにおけるカプトプリルの効果

	カプトプリル群と プラセボ群の差	95%信頼区間		p 値
平均の比較	−6.52	−14.25	1.20	0.12
変化量の比較	−7.95	−16.39	0.48	0.09
共分散分析	−7.18	−12.99	−1.37	0.03

演 習 問 題

〈最小 2 乗推定量の導出〉

問 1　対数尤度関数

$$l(\boldsymbol{\beta},\sigma) = \sum_{i=1}^{N} \frac{-1}{2}\left(\frac{y_i - \boldsymbol{X}_i\boldsymbol{\beta}}{\sigma}\right)^2$$

から，分散 σ^2 が既知のときの $\boldsymbol{\beta}$ のスコア関数と Fisher 情報行列を求めよ．そして，スコア方程式の解（最尤推定量）$\widehat{\boldsymbol{\beta}}$ を導出せよ．また，Fisher 情報行列を用いて，$\widehat{\boldsymbol{\beta}}$ の分散が $\sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ であることを示せ．

12 2値データの回帰モデル

この章では，2 値データの回帰モデルについて解説する．プロビット回帰，ロジスティック回帰，積 2 項分布モデルとその応用（用量反応関数の推定，判別分析，ランダム化臨床試験の解析）について述べる．

一般化線型モデルでは，リンク関数の選択によって，効果の指標を指定できる．2 値データの解析では，効果の指標としてリスク差，リスク比，オッズ比が用いられるが，それぞれ恒等リンク，対数リンク，ロジットリンクに対応している．

キーワード ROC 曲線，オッズ比，完全分離，C 統計量，積 2 項分布モデル，2 項分布，判別分析，プロビット回帰，ロジスティック回帰，ロジット関数，用量反応関係，リスク差，リスク比

事　例 英国 ECMO 試験，糸球体濾過率研究，6 都市研究

12.1 モデルの構造

アウトカム Y_i が，0 または 1 で表される個人の反応の有無を表す 2 値変数であり，$\Pr(Y_i = 1) = \pi_i$ という確率で独立に分布しているとする．このとき N 人の対象者の同時分布は，次のように表現できるから，指数型分布族のひとつであることがわかる．

$$\prod_{i=1}^{N} \pi_i{}^{y_i} (1-\pi_i)^{1-y_i} = \exp\left[\sum_{i=1}^{N} y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \sum_{i=1}^{N} \log(1-\pi_i)\right]$$

これを一般化線型モデルとして扱う場合には，確率パラメータ π_i と共変量 X_1, X_2,…,X_p の関係は，リンク関数とデザイン行列を用いて

$$g(\pi_i) = \boldsymbol{X}_i\boldsymbol{\beta}$$

と表される．

12.2 推　　定

このモデルはパラメトリックモデルの一種だから，10 章で述べた最尤法を利用して推定することができる．ただしこのケースではかならずしもデータから最尤推定量が求まるわけではない．計算上の問題として完全分離（complete separation）という現象が知られている．完全分離とは，$\boldsymbol{X}_i\widehat{\boldsymbol{\beta}}$ を計算したとき，その値によって，すべての対象者を 100%の精度で 0 または 1 に判別できてしまう状況のことをいう．完全分離が生じたとき，データにパラメータを推定するための情報が不足していることを意味しているから，そのままでは一般化線型モデルを当てはめることはできない．共変量の数の削減やペナルティ付き尤度の利用など，なんらかの対処が必要である．

12.3 用量反応関係

歴史的にバイオアッセイの分野では，2 値データの回帰モデルが用いられてきた．この分野では，毒性物質のいくつかの用量について，動物の死亡割合が調べられる．そのときの目的は，死亡確率 π を用量 X の関数とみなした用量反応関数を推定することである．用量反応関数は，用量 X を動かしたときに [0, 1] までの区間に制限されなければならない．この条件を満たすためのテクニックとして，なんらかの確率密度関数 $p(y)$ を用いて

$$\pi = \int_{-\infty}^{X} p(y)dy$$

というモデルを用いることがある．初期のバイオアッセイで用いられたのは，$p(y)$ に正規分布を用いたプロビット関数

$$\pi = \Phi\left(\frac{X-\mu}{\sigma}\right)$$

である（Φ は正規分布の分布関数を表す）．このモデルは，$\beta_0 = -\mu/\sigma$，$\beta_1 = 1/\sigma$ とおけば

$$\Phi^{-1}(\pi) = \beta_0 + \beta_1 X$$

だから，プロビット回帰は，リンク関数に $g(x) = \Phi^{-1}(x)$ を用いた一般化線型モデルであることがわかる．

プロビット関数に代わって広く用いられており，ほとんど同じ関数形をしているのがロジット関数

$$g(x) = \log\left(\frac{x}{1-x}\right)$$

である．このときのモデルはロジスティック回帰と呼ばれ，用量反応関数に

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X$$

を仮定していることになる．ロジット関数は 2 項分布の正準リンク関数である．

ロジット関数の例を図 12-1 に示す．これは，ロジット関数の $\exp(\beta_1)$ の値を 5，10，20，200 と設定して，X を 0 から 1 までの範囲で変化させたプロットである．ロジット関数はこのように S 字型の曲線を表していて，曲線の傾きは X の係数によって決まる．

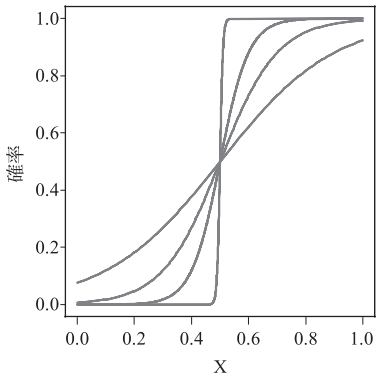

図 12-1 ロジット関数で結びついた 2 変数の関係の例
1SD あたりのオッズ比を 5，10，20，200 と動かしたもの

12.4 判別分析

■ 12.4.1 ロジット関数の導出

ロジット関数は，用量反応関係を調べるときだけでなく，2 値判別でも用いられる．2 つのグループから構成される母集団があって，それぞれのグループを陰性（$Y_i = 0$）と陽性（$Y_i = 1$）で表す．ある個人から連続データ X_i が得られているとき，その値からその個人が 2 つのグループのどちらに属するかを判

別するというのが 2 値判別の問題である．

陰性と陽性それぞれの構成割合を $\Pr(Y_i = 0) = 1 - \pi$ と $\Pr(Y_i = 1) = \pi$ とする．また，X_i の条件付確率密度関数を $p(x|Y_i)$，その比を $l(x) = \log[p(x|Y_i = 1)/p(x|Y_i = 0)]$ とする．この比のことを尤度比と呼ぶことがある．Bayes の定理から

$$\frac{\Pr(Y_i = 1|X_i = x)}{\Pr(Y_i = 0|X_i = x)} = \frac{p(x|Y_i = 1)}{p(x|Y_i = 0)} \frac{\pi}{1 - \pi}$$

が成り立つから，その対数をとれば

$$\log \left[\frac{\Pr(Y_i = 1|X_i = x)}{\Pr(Y_i = 0|X_i = x)} \right] = \log \left(\frac{\pi}{1 - \pi} \right) + l(x)$$

と表すことができる．ここで X_i が正規分布に従うと仮定する．このとき，簡単な計算から，正規密度関数の比は $l(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ という形式になることが示される．ここで，$\Pr(Y_i = 1|X_i = x) = \pi_i$ とおくと，正規分布の仮定は

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = \left[\log \left(\frac{\pi}{1 - \pi} \right) + \beta_0 \right] + \beta_1 X_i + \beta_2 X_i^2$$

というロジスティック回帰を当てはめていることと等しいことがわかる．さらに，正規分布の分散が共通のとき，$l(x) = \beta_0 + \beta_1 x$ というように 2 次の項が消える．これは

$$\log \left(\frac{\pi_i}{1 - \pi_i} \right) = \left[\log \left(\frac{\pi}{1 - \pi} \right) + \beta_0 \right] + \beta_1 X_i$$

という 1 次のロジスティック回帰に帰着する．この手続きは，事前確率 π を，データ $l(X_i)$ を用いて事後確率 π_i に更新するという Bayes 推測そのものである．

ここで述べた結果は，ロジットリンクと条件付確率密度関数 $p(x|Y_i)$ の比の関係を表している．もし，2 つのグループからデータ X_i を集められれば，確率密度関数 $p(x|Y_i)$ を得ることができる．そして X_i の分布が正規分布に近ければ，1 次または 2 次のロジスティック回帰を当てはめ π_i を推定し，その値により陰性と陽性を分類することで，個人 i がどちらのグループに属するかを判別できる．判別分析において，ロジットリンクが自然に導かれるわけである．

ロジスティック回帰は，プロビット回帰など他の 2 値データの回帰モデルを過去のものにした．実際，X_i の分布が単峰性であれば，ロジスティック回帰は実用上問題ないくらいデータへの当てはまりがよい．さらに他のリンク関数に比べて，正準リンクであるロジット関数は計算が安定する．

■ 12.4.2 ROC 曲線と C 統計量

感度と特異度

一般に，連続データ X_i によって陰性と陽性を判別する能力（descrimination）は，感度・特異度によって表される．X_i の値をカットオフ値 c と比較することで，2 つのグループの判別を行うとしよう．仮に X_i が高いほど $Y_i = 1$ の確率が高いという方向性があり，$X_i \geq c$ のとき陽性と判断する．このときカットオフ値 c を用いたときの感度と特異度は

$$\Pr(X_i \geq c | Y_i = 1)$$

$$\Pr(X_i < c | Y_i = 0)$$

と定義される．

陽性者・陰性者の人数だけでなく，感度・特異度も c を動かすことで変化する．カットオフ値 c が高いほど，陽性者の人数は少なくなるが，感度は向上することが普通である．感度と特異度にはトレードオフの関係がある．カットオフ値を上昇させることで，特異度を高められるが，一方で感度は減少してしまう．カットオフ値を最低値にすることで（全員を陽性と判断することで），感度は 100%にできるが特異度は 0%になる．

もし，判別に用いるデータが $\boldsymbol{X}_i = (1, X_{i1}, X_{i2}, \ldots, X_{ip})^T$ というように複数あったとしても，一般化線型モデルを仮定することで，総合的な判別精度を評価することができる．たとえば

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \boldsymbol{X}_i\boldsymbol{\beta}$$

というロジスティック回帰を仮定したとしたら，感度・特異度は

$$\Pr(\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} \geq c | Y_i = 1)$$

$$\Pr(\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} < c | Y_i = 0)$$

によって計算される．上のモデルは主効果のみを含めたが，もちろん 2 次以上の項や交互作用をモデルに加えることもできる．

ROC 曲線

カットオフ値を変化させたときの感度と 1 – 特異度 を，それぞれ縦軸と横軸にプロットしたものが receiver-operator-characteristic（ROC）曲線である．ROC 曲線は，45 度の直線から離れるほど（図の左上に近づくほど），感度・特異度が

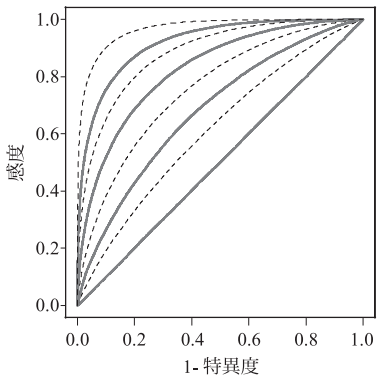

図 12-2 ROC 曲線の例
1SD あたりのオッズ比を 1，1.5，2，3，5，10，20，200 と動かしたもの

高いことを意味する．ROC 曲線の曲線下面積は C 統計量と呼ばれ，$C = 1$ は感度・特異度 100%が達成できたとき，$C = 0.5$ は対象者をランダムに分類したときに対応する．ROC 曲線と C 統計量は，データ X_i の値そのものではなく，X_i と c の大小関係の情報しか利用していないことに注意しよう．そのためこれらの指標は，X_i の順位情報だけに依存する．実際，C 統計量は順位統計量の一種（Wilcoxon–Mann–Whitney［ウィルコクソン・マン・ホイットニー］U 統計量）を 0 から 1 の範囲にスケールを直したものである．C 統計量の信頼区間は，この関係を利用して構成される．

■ 12.4.3 ROC 曲線とオッズ比の関係

図 12-2 は，ロジスティック回帰

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_i$$

の下で，1SD あたりのオッズ比を 1，1.5，2，3，5，10，20，200 と動かしたときの ROC 曲線である[*1)]．図によると，C 統計量が 0.9 を超えるのはオッズ比が 20 倍のときである（$C = 0.93$）．これは $Y_i = 0$ と $Y_i = 1$ の平均の差が 1.46SD である状況に相当する．このように，ROC 曲線と C 統計量は，（オッズ比 1.5 ~ 5 程度の）比較的弱い関連性に対して鋭敏な性能評価指標ではない[*2)]．

[*1)] 1SD あたりのオッズ比は，X_i の SD が 1 のとき $\exp(\beta_1)$ のことである．

[*2)] 疫学研究では，20 倍のオッズ比をみることはまれであり，たとえば循環器リスク因子は，心血管疾患を 1.5 ~ 5 倍増やす程度の影響に過ぎない．そのため，疫学的に強いリスク因子であって

医学で疾患の有無（有病）を判別するときは，症状や病変などなんらかの兆候が生じていることが前提である．よい診断法とは，診断時に存在する兆候を見落とすことなく検出することであり，C 統計量でいえば 0.9 を超えるような精度が求められる．一方でリスク因子を探索する疫学研究では，数年後に疾患が発生するかどうかの予測を試みる．

疾患発生は，予測時点のリスク因子の有無だけではなく，さまざまな要因が関係するだろう．そのため疫学研究でオッズ比 10 ~ 20 程度のリスク因子を見出すことは難しい．図 12-2 は，診断研究と疫学研究が，定量的に異なる問題を使っていることを示している．

■ 12.4.4 事例：クレアチニンによる慢性腎臓病の診断精度の評価

表 12-1 は，11 章の糸球体濾過率研究において，血漿クレアチニン濃度によって慢性腎臓病を判別できるかを調べるためのデータである．

まず，アウトカムを糸球体濾過率 60 mL/min 未満によって診断した慢性腎臓病の有無としたとき，血漿クレアチニン濃度との関係はどのようなものであろ

表 12-1 糸球体濾過率研究データ

対象	慢性腎臓病（糸球体濾過率 60 mL/min 未満）	クレアチニン（mg/dL）	対象	慢性腎臓病（糸球体濾過率 60 mL/min 未満）	クレアチニン（mg/dL）
1	なし	0.85	17	あり	1.83
2	あり	0.99	18	あり	1.98
3	なし	1.13	19	あり	2.03
4	なし	1.13	20	あり	2.09
5	なし	1.13	21	あり	2.77
6	なし	1.13	22	あり	2.96
7	なし	1.13	23	あり	3.11
8	なし	1.27	24	あり	3.96
9	あり	1.41	25	あり	4.69
10	あり	1.47	26	あり	4.8
11	なし	1.47	27	あり	5.93
12	あり	1.56	28	あり	5.93
13	あり	1.69	29	あり	5.93
14	なし	1.7	30	あり	7.79
15	あり	1.75	31	あり	11.02
16	あり	1.75			

も，ROC 曲線と C 統計量を求めるとがっかりすることがある．C 統計量は 0.6 ~ 0.7 程度にしかならないからである．

うか．これはアウトカムを慢性腎臓病の有無，共変量を血漿クレアチニン濃度として

$$\log(\text{ODDS of CKD}) = \text{INTERCEPT} + \text{INVERSE of CREATININE}$$

というロジスティック回帰を当てはめることで調べられる．図 12-3 はこのモデルから推定されたロジット関数をプロットしたものである．クレアチニン逆数と慢性腎臓病の確率との関係は，ほぼ直線的であった（図 12-3 左）．クレアチニン逆数 1SD あたりのオッズ比は 19.0（95%信頼区間 2.2 ~ 163.5，$p < 0.01$）と推定された．

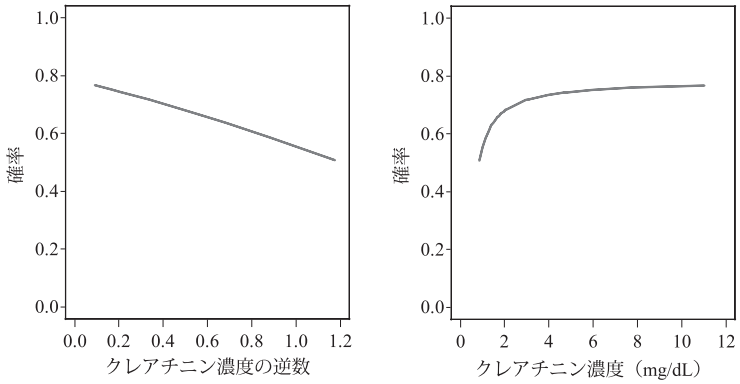

図 12-3 糸球体濾過率データにおけるクレアチニン逆数と慢性腎臓病の確率との関係（左）とその横軸を血漿クレアチニン濃度に変えたもの

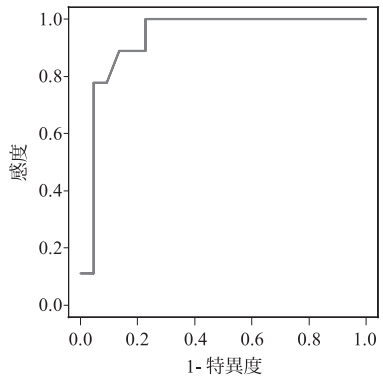

図 12-4 糸球体濾過率データにおける血漿クレアチニン濃度による慢性腎臓病の判別精度を表す ROC 曲線

図 12-4 は，血漿クレアチニン濃度による慢性腎臓病の判別精度を表す ROC 曲線である．C 統計量は 0.93（95%信頼区間 0.84 ~ 1.00）であった．

12.5 2 × 2 表と積 2 項分布

■ 12.5.1 リスク差，リスク比，オッズ比

臨床試験や疫学研究では，2 群の 2 値アウトカムを比較するとき，データを 2 × 2 表にまとめることが多い．そして 2 群の差を表す効果の指標として，リスク差，リスク比，オッズ比などが用いられる．回帰モデルを用いて，これらの指標を表現してみよう．

対象者 i に試験治療を用いたかどうかを A_i（$A_i = 0$ ならコントロール治療，$A_i = 1$ なら試験治療群），アウトカムを Y_i（$Y_i = 0$ ならイベントなし，$Y_i = 1$ ならイベントあり）で表す．N，N_0，N_1 をそれぞれ全体，コントロール群，試験治療群の人数とする．また，S，S_0，S_1 をそれぞれ全体，コントロール群，試験治療群のイベント数とする．表 12-2 はこのデータの記法を示したものである．2 群の違いは，

$$\text{Risk difference} = \frac{S_1}{N_1} - \frac{S_0}{N_0}$$

$$\text{Risk ratio} = \frac{S_1}{N_1} \div \frac{S_0}{N_0}$$

$$\text{Odds ratio} = \frac{S_1}{N_1 - S_1} \div \frac{S_0}{N_0 - S_0}$$

という指標を求めることで調べられる．

表 12-2 2 × 2 表の記法

	試験治療（$A = 1$）	コントロール（$A = 0$）	合計
イベントなし（$Y = 0$）	$N_1 - S_1$	$N_0 - S_0$	
イベントあり（$Y = 1$）	S_1	S_0	S
合計	N_1	N_0	N

ここで，S_0 と S_1 は独立な 2 項分布に従うと仮定する．これを積 2 項分布モデル（product binomial model）という．コントロール群のリスクを π^0，試験治療群のリスクを π^1 と表すと，対数尤度関数は

$$l(\boldsymbol{\pi}) = S_0 \log(\pi^0) + (N_0 - S_0)\log(1-\pi^0) + S_1 \log(\pi^1) + (N_1 - S_1)\log(1-\pi^1)$$

のように，2 項尤度の和の形で表すことができる．これまで述べてきたように，対数尤度関数を最大化することで，π^0 と π^1 の最尤推定量を求めることができる．ただし，ここで関心があるのは，π^0 と π^1 自体ではなく，効果の指標である．リンク関数を適切に選ぶことによって，リスク差（$\pi^1 - \pi^0$），リスク比（π^1/π^0），オッズ比（$\pi^1/(1-\pi^1)/[\pi^0/(1-\pi^0)]$）をそれぞれ指定できる．

リンク関数のうちもっとも単純な恒等リンク $g(x) = x$ は

$$\pi^a = \beta_0 + \beta_1 a$$

という関係を意味する．この式に $a = 0$ を代入すれば，$\beta_0 = \pi^0$ という対応が明らかになる．また，回帰係数 β_1 はリスク差そのものである．

次に，対数リンク $g(x) = \log(x)$ を用いれば

$$\pi^a = \exp(\beta_0 + \beta_1 a)$$

という対数線型モデルとなる．ここで，$\exp(\beta_1)$ はリスク比に対応する．

最後にロジットリンク $g(x) = \log[x/(1-x)]$ は

$$\frac{\pi^a}{1-\pi^a} = \exp(\beta_0 + \beta_1 a)$$

というように，オッズについての対数線型モデルである．$\exp(\beta_1)$ はオッズ比になる．

■ 12.5.2　対数尤度

この 3 つのリンク関数は，同一の積 2 項分布において，パラメータ表現を変えたものである．さらに，どのリンク関数であっても，確率パラメータ (π^0, π^1) と回帰係数 (β_0, β_1) の値には 1 対 1 の対応関係がある．したがって，回帰係数 (β_0, β_1) の最尤推定量は，積 2 項分布の対数尤度関数

$$l(\boldsymbol{\pi}) = S_0 \log(\pi^0) + (N_0 - S_0)\log(1-\pi^0) + S_1 \log(\pi^1) + (N_1 - S_1)\log(1-\pi^1)$$

をリンク関数を介して最大にする値である．

ロジットリンクの場合でこれを確かめてみよう．ロジットリンクは

$$\pi^0 = \frac{1}{1 + [\exp(\beta_0)]^{-1}}$$

$$\pi^1 = \frac{1}{1 + [\exp(\beta_0 + \beta_1)]^{-1}}$$

という対応関係を意味する．これを対数尤度関数に代入して

$$l(\boldsymbol{\beta}) = S_0\beta_0 + S_1\beta_1 - N_0 \log[1 + \exp(\beta_0)] - N_1 \log[1 + \exp(\beta_0 + \beta_1)]$$

という $\boldsymbol{\beta}$ の対数尤度関数が得られる．

$l(\boldsymbol{\pi})$ は 2 つの 2 項尤度の和なので，この尤度に基づく最尤推定量はもともとの 2 項分布の割合に帰着する．つまり，一方のパラメータは他方のパラメータの推定に影響しない．ところが $l(\boldsymbol{\beta})$ には，β_0 と β_1 の両方を含む項がある．これは観測情報行列の非対角要素がゼロではないことを意味する．図 12-5 は，1996 年に英国で行われた ECMO 臨床試験（UK Collaborative ECMO Trial Group 1996）から得られた $l(\boldsymbol{\pi})$ と $l(\boldsymbol{\beta})$ を，三次元プロットと等高線プロットによって図示したものである．等高線上において，β_0 と β_1 は楕円に近い関係にある

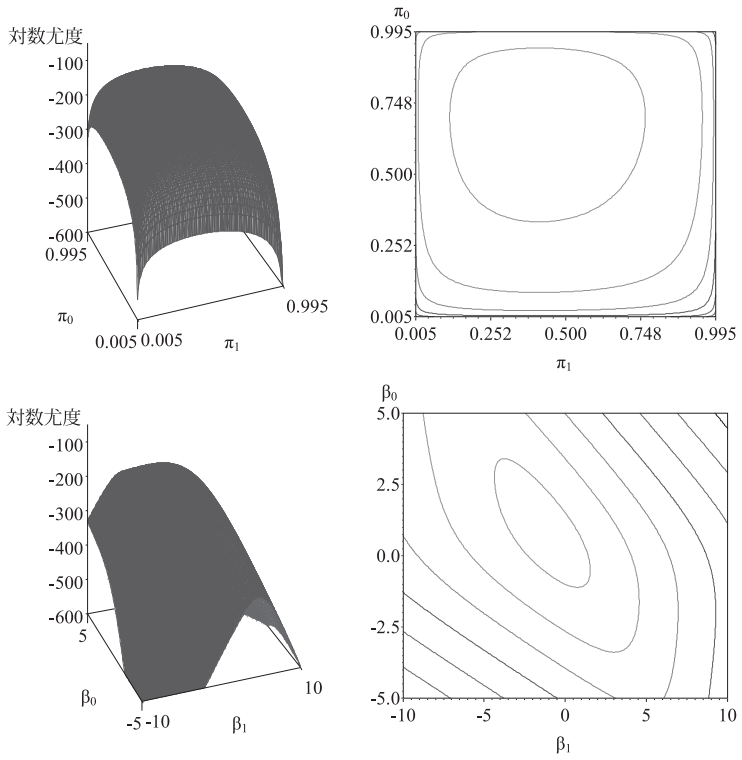

図 12-5 積 2 項分布モデルの対数尤度関数
上は確率パラメータ $\boldsymbol{\pi}^0$，$\boldsymbol{\pi}^1$ の 3 次元プロットと等高線プロット，
下は回帰係数 $\boldsymbol{\beta}_0$，$\boldsymbol{\beta}_1$ の 3 次元プロットと等高線プロット

ことがわかる．これは $\widehat{\beta_0}$ と $\widehat{\beta_1}$ の間に負の相関があることを示唆する．

■ 12.5.3　最尤推定量

次に最尤推定量とその漸近分布を導出してみよう．$\boldsymbol{\pi} = (\pi^0, \pi^1)^T$ の最尤推定量は，スコア方程式

$$\boldsymbol{U}(\boldsymbol{\pi}) = \begin{bmatrix} \frac{S_0 - N^0\pi^0}{\pi^0(1-\pi^0)} \\ \frac{S_1 - N^1\pi^1}{\pi^1(1-\pi^1)} \end{bmatrix} = \boldsymbol{0}$$

の解である．これを求めるとそれぞれの群の割合

$$\widehat{\boldsymbol{\pi}} = \begin{pmatrix} \frac{S_0}{N_0} \\ \frac{S_1}{N_1} \end{pmatrix}$$

になる．ここから対数オッズ比は

$$\log\left[\frac{\widehat{\pi}^1/(1-\widehat{\pi}^1)}{\widehat{\pi}^0/(1-\widehat{\pi}^0)}\right] = \log\left[\frac{S_1(N_0 - S_0)}{S_0(N_1 - S_1)}\right]$$

と計算される．

一方，ロジットリンクを用いて

$$\pi^0 = \frac{1}{1 + [\exp(\beta_0)]^{-1}}$$

$$\pi^1 = \frac{1}{1 + [\exp(\beta_0 + \beta_1)]^{-1}}$$

というパラメータ表現を用いるなら，これを対数尤度関数に代入して

$$l(\boldsymbol{\beta}) = (S_0 + S_1)\beta_0 + S_1\beta_1 - N_0 \log[1 + \exp(\beta_0)] - N_1 \log[1 + \exp(\beta_0 + \beta_1)]$$

という $\boldsymbol{\beta}$ の対数尤度関数が得られる．スコア方程式は

$$\boldsymbol{U}(\boldsymbol{\beta}) = \begin{bmatrix} S_0 + S_1 - \frac{N_0}{1+[\exp(\beta_0)]^{-1}} - \frac{N_1}{1+[\exp(\beta_0+\beta_1)]^{-1}} \\ S_1 - \frac{N_1}{1+[\exp(\beta_0+\beta_1)]^{-1}} \end{bmatrix} = \boldsymbol{0}$$

であり，その解として

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \log\left(\frac{S_0}{N_0 - S_0}\right) \\ \log\left[\frac{S_1(N_0 - S_0)}{S_0(N_1 - S_1)}\right] \end{bmatrix}$$

が導かれる．最尤推定量を比較すれば，$U(\boldsymbol{\pi}) = 0$ と $U(\boldsymbol{\beta}) = 0$ は，同一の対数オッズ比の値を導いているから，不変性が成り立っていることがわかる．

一方で，両者から得られる推定量の漸近分布はそれぞれ

$$\widehat{\boldsymbol{\pi}} \sim N\left[\begin{pmatrix}\pi^0\\ \pi^1\end{pmatrix},\ \begin{pmatrix}\frac{N^0}{\pi^0(1-\pi^0)} & 0\\ 0 & \frac{N^1}{\pi^1(1-\pi^1)}\end{pmatrix}\right]$$

と

$$\widehat{\boldsymbol{\beta}} \sim N\left[\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix},\ \begin{pmatrix}\frac{1}{N_0\pi^0(1-\pi^0)} & \frac{-1}{N_0\pi^0(1-\pi^0)}\\ \frac{-1}{N_0\pi^0(1-\pi^0)} & \frac{1}{N_0\pi^0(1-\pi^0)}+\frac{1}{N_1\pi^1(1-\pi^1)}\end{pmatrix}\right]$$

である．ここで注目してほしいのは，Wald 信頼区間の構成方法は，$\widehat{\boldsymbol{\pi}}$ の正規近似に基づくものと $\widehat{\boldsymbol{\beta}}$ を正規近似したものの 2 通りがあり得るということである．π^0 と π^1 は，0 から 1 までの値しかとらないという制約があるから，$\widehat{\boldsymbol{\beta}}$ の方が正規近似への当てはまりがよい．このように，最尤法の点推定値はパラメータ変換について不変だが，Wald 信頼区間はそうではない．

疫学の分野では歴史的にロジスティック回帰がよく用いられてきた．その主な理由は小標本特性がよく，ケース・コントロール研究という一部の対象者しかサンプリングされない場合も，切片項を除いて一致推定量が得られるためである．しかし，それ以外の状況では，恒等リンク・対数リンクの方が，効果の指標を解釈しやすいという点で好ましい．

12.6 事例：2 値アウトカムの臨床試験の解析 4

■ 12.6.1 リスク差，リスク比，オッズ比の推定

表 12-3 に英国 ECMO 試験のデータを示す（UK Collaborative ECMO Trial Group 1996）．この試験は，典型的なランダム化臨床試験であり，2×2 表のための標準的な手法で解析できる．表 12-4 に恒等リンク・対数リンク・ロジットリンクを用いた解析結果を示す．ECMO の効果を解釈するうえで，リスク差・リスク比のどちらも有用な情報だから，この場合は両方を報告すべきである．また，Fisher の正確検定を用いても結果は $p < 0.01$ であった．

表 12-3 英国 ECMO 試験データ

	ECMO	従来療法
生存	65	38
死亡	28	54
合計	93	92
死亡割合	30.1%	58.7%

表 12-4　英国 ECMO 試験データにおける従来療法と比べた ECMO の効果

	推定値	95%信頼区間		p 値
リスク差	−28.6%	−42.3	−14.9	< 0.01
リスク比	0.59	0.45	0.78	< 0.01
オッズ比	0.30	0.17	0.56	< 0.01

■ 12.6.2　事例から得られた教訓

ECMO の事例は統計学では有名で，ランダム化の倫理性，アウトカム適応的ランダム化の特徴，頻度論と Bayes 流の違いなどさまざまな議論がなされた．この事例から学ぶべき教訓のひとつは，因果推論が妥当であるためには，統計解析以上に研究計画が大切ということである．英国 ECMO 試験のように，ランダム化とサンプルサイズ計算が適切になされていれば，単純な統計手法で因果効果を正しく推定することができる．それ以前のハーバード試験・ミシガン試験は，今の目でみるとやはり研究としての質が低く，ECMO を用いるべきかという問題について，医学界のコンセンサスを得られるような答えは得られなかった．