一般化線形モデル（区間推定，仮説検定，正規線形モデル）

1 はじめに

一般化線形モデルにおける，**信頼区間**と**仮説検定**の導出およびその利用について述べる。信頼区間は**区間推定**としても知られ，信頼区間がある値を含むかどうかが検定結果，その幅が検定精度（検出力）の情報も与えるため，仮説検定より有用であるとみなされるになっている。信頼区間の幅は，仮説検定の検出力よりも概念的に単純で理解しやすい。

統計モデルの枠組みにおける仮説検定は，2 つの関連するモデルがデータにいかによく適合するかを比べることによってなされる。一般化線形モデルにおいては，2 つのモデルは同一の確率分布と同一の連結関数をもち，一方のモデルの線形成分は他方よりも多くのパラメータをもつとする。単純なモデルは帰無仮説 H_0 に対応し，他方のより一般的なモデルの特別な場合でなければならない。もし，単純なモデルが，より一般的なモデルと同程度データに適合するのであれば，倹約の観点から単純なモデルが選択され，H_0 が採択される。もし，より一般的なモデルの方が有意に適合すれば，H_0 を棄却し，より一般的なモデルに対応する対立仮説 H_1 を採択する。このような比較をするために，モデルがいかによくデータに適合するかを評価する指標が必要となる。モデルがデータにどの程度適合しているかを測る指標である**適合度統計量**は，尤度関数の最大値，対数尤度関数の最大値，平方和基準の最小値，残差に基づく複合型の統計量などを基礎として構築される。その手順は，次のように要約される。

1. H_0 に対応するモデル M_0 とより一般的なモデル M_1 を規定する。ただし，M_0 は M_1 の特別な場合とする。
2. M_0 をデータに適用し，適合度を測る統計量 G_0 を計算する。M_1 をデータに適用し，適合度を測る統計量 G_1 を計算する。
3. 適合度の改善を計算する。$G_1 - G_0$ を用いることが多いが，G_1/G_0 を用いる場合もある。
4. 対立仮説 $G_1 \neq G_0$ に対して，帰無仮説 $G_1 = G_0$ を検定するため，$G_1 - G_0$ の標本分布を用いる。
5. もし帰無仮説 $G_1 - G_0$ が棄却されなければ，M_0 がより良いモデルとなる。一方，帰無仮説 $G_1 - G_0$ が棄却されれば，M_1 がより良いモデルとなる。

区間推定と仮説検定のどちらの推測でも，統計量の標本分布が必要となる。つまり，信頼区間を計算するためには推定量の標本分布が必要になり，仮説検定をするためには適合度を測る統計量の標本分布が必要となる。

もし応答変数が正規分布に従うのであれば，その推測に用いられる標本分布は多くの場合正確に求まる。その他の分布の場合は，中心極限定理に基づく大標本の漸近的な結果を用いる必要がある。これらの結果を厳密に導出するためには，様々な正則条件を詳細に検討する必要がある。指数分布族に属する分布からの独立な観測値の場合，つまり一般化線形モデルの場合には，必要な条件を満足している。統計学特論では，標本分布の導出過程のみを考え，細かい点については取り扱わない。

基本的な考え方は，S が関心のある統計量であるならば，適当な条件の下で，漸近的に

$$\frac{S-E[S]}{\sqrt{V[S]}} \sim N(0,1)$$

または，

$$\frac{(S-E[S])^2}{V[S]} \sim \chi^2(1)$$

が成り立つことを示し，これを用いることである。ここで，$E[S]$ と $V[S]$ はそれぞれ S の漸近的な期待値と分散である。

関心のある統計量がベクトル

$$\boldsymbol{s} = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_p \end{pmatrix}$$

の場合には，漸近的な期待値 $E[\boldsymbol{s}]$ と分散共分散行列 $V[\boldsymbol{s}]$（ただし，$V[\boldsymbol{s}]$ は正則で逆行列 $(V[\boldsymbol{s}])^{-1}$）に対して，漸近的に

$$[\boldsymbol{s}-E[\boldsymbol{s}]]^\top V[\boldsymbol{s}])^{-1}[\boldsymbol{s}-E[\boldsymbol{s}]] \sim \chi^2(p)$$

が成り立つことを示し，これを利用する。

2　スコア統計量の標本分布

$Y_1,\ldots,Y_n$ は，パラメータ $\boldsymbol{\beta}$ をもつ一般化線形モデルにおける互いに独立な確率変数であり，次式が成り立つとする。

$$E[Y_i]=\mu_i, \quad g(\mu_i)=\boldsymbol{x}_i^\top\boldsymbol{\beta}=\eta_i \quad (i=1,\ldots,n)$$

スコア統計量は，次式となる。

$$U_j=\sum_{i=1}^{n}\left[\frac{Y_i-\mu_i}{V[Y_i]}\cdot x_{ji}\cdot\frac{\partial\mu_i}{\partial\eta_i}\right] \quad (j=0,1,\ldots,p)$$

ここで，すべての i について $E[Y_i]=\mu_i$ であることから $E[U_j]=0$ $(j=0,1,\ldots,p)$ となる。これは，前回の講義で示した指数分布族の一般的な結果に一致する。スコア統計量 U_j $(j=0,1,\ldots,p)$ の分散共分散行列は情報行列 $\boldsymbol{\mathfrak{I}}$ とよばれる。前回の講義で示した通り，情報行列 $\boldsymbol{\mathfrak{I}}$ の (k,l) 要素は次のように表される。

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

単一のパラメータ β のみがモデルに含まれる場合，前回の講義で示した通り，$E[U]=0$ かつ $V[U]=\mathfrak{I}$ である。したがって，スコア統計量は次の漸近標本分布をもつことがわかる（詳細は，補足 1 を参照）。

$$\frac{U}{\sqrt{\mathfrak{I}}} \sim N(0,1), \quad \frac{U^2}{\mathfrak{I}} \sim \chi^2(1)$$

パラメータベクトル $\boldsymbol{\beta}$ がモデルに含まれる場合，スコアベクトル $\boldsymbol{U}=\dfrac{\partial L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}$ は，漸近的に多変量正規分布 $N_{p+1}(\boldsymbol{0},\boldsymbol{\mathfrak{I}})$ に従う。大標本においては

$$\boldsymbol{U}^\top\boldsymbol{\mathfrak{I}}^{-1}\boldsymbol{U} \sim \chi^2(p+1)$$

が成り立つ。

補足 1

対数尤度関数を $l = \sum_{i=1}^{n} l_i$ とする。ただし，$l_i = \log f(Y_i|\beta)$ である。スコア統計量を $U = \sum_{i=1}^{n} U_i$ とする。ただし，$U_i = \dfrac{\partial l_i}{\partial \beta}$ である。ここで，l_i と U_i は確率変数 Y_i の関数である。$Y_1, \ldots, Y_n$ は互いに独立とすると，$U_1, \ldots, U_n$ は互いに独立で，かつ同一の分布に従う確率変数となる。スコア統計量 U はそのような確率変数の和になっていることから，中心極限定理より漸近的に正規分布に従うことがわかる。さらに，$E[U] = 0$ かつ $V[U] = \mathfrak{I}$ であることから，漸近的に $U \sim N(0, \mathfrak{I})$ となる。パラメータがベクトルの場合も同様に，$\boldsymbol{U} \sim N_{p+1}(\boldsymbol{0}, \boldsymbol{\mathfrak{I}})$ となる。

問 4

$Y_1, \ldots, Y_n$ を独立に同一分布に従う確率変数で，$Y_i \sim N(\mu, \sigma^2)$ とする。ただし，関心のあるパラメータは μ であり，σ^2 は既知の定数とする。このとき，対数尤度関数は次式となる。

$$l = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 - n \log\left(\sqrt{2\pi\sigma^2}\right)$$

スコア統計量は，次式となる。

$$U = \frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \mu) = \frac{n}{\sigma^2} (\bar{Y} - \mu)$$

ただし，$\bar{Y} = \dfrac{1}{n} \sum_{i=1}^{n} Y_i$ とする。このとき，以下の各問に答えよ。

(1) 方程式 $U = 0$ を解くことにより，最尤推定量 $\hat{\mu}$ を求めよ。
(2) $E[U] = 0$ となることを示せ。
(3) $V[U]$ を求めよ。
(4) $\dfrac{U}{\sqrt{\mathfrak{I}}}$ は，（漸近ではなく）正確に $N(0, 1)$ に従う。この理由を述べよ。

3 最尤推定量の標本分布

単一のパラメータ β の対数尤度関数に対して，推定値 $\hat{\beta}$ の近くでテイラー展開すると，最初の 3 項は次のようになる。

$$l(\beta) = l(\hat{\beta}) + (\beta - \hat{\beta}) U(\hat{\beta}) + \frac{1}{2} (\beta - \hat{\beta})^2 U'(\hat{\beta})$$

ここで，$U(\hat{\beta}) = \dfrac{\partial l}{\partial \beta}$ は，$\beta = \hat{\beta}$ に対して計算されたスコア関数である。$U' = \dfrac{\partial^2 l}{\partial \beta^2}$ がその期待値 $E[U'] = -\mathfrak{I}$ によって近似できるならば，次の近似式を得る。

$$l(\beta) = l(\hat{\beta}) + (\beta - \hat{\beta})U(\hat{\beta}) - \frac{1}{2}(\beta - \hat{\beta})^2 \mathfrak{I}(\hat{\beta})$$

ここで，$\mathfrak{I}(\hat{\beta})$ は，$\beta = \hat{\beta}$ に対して計算された情報量である。

単一のパラメータのスコア関数に対して，推定値 $\hat{\beta}$ の近くでテイラー展開すると，最初の 2 項は次のようになる。

$$U(\beta) = U(\hat{\beta}) + (\beta - \hat{\beta})U'(\hat{\beta})$$

U' が $E[U'] = \mathfrak{I}$ で近似できるとき，次式が得られる。

$$U(\beta) = U(\hat{\beta}) - (\beta - \hat{\beta})\mathfrak{I}(\hat{\beta})$$

同様に，パラメータベクトル $\boldsymbol{\beta}$ の対数尤度関数に対して，推定値 $\hat{\boldsymbol{\beta}}$ の近くでテイラー展開すると，最初の 3 項は次のようになる。

$$l(\boldsymbol{\beta}) = l(\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \boldsymbol{U}(\hat{\boldsymbol{\beta}}) - \frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \boldsymbol{\mathfrak{I}}(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

パラメータベクトル $\boldsymbol{\beta}$ のスコア関数の近似式は次のようになる。

$$\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{U}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\mathfrak{I}}(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

この式は，最尤推定量 $\hat{\boldsymbol{\beta}}$ の標本分布を得るのに用いることができる。最尤推定量の定義より，$\hat{\boldsymbol{\beta}}$ は $l(\boldsymbol{\beta})$ を最大にする推定量であることから，$\boldsymbol{U}(\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$ を満たす必要がある。したがって，次式を得る。

$$\boldsymbol{U}(\boldsymbol{\beta}) = -\boldsymbol{\mathfrak{I}}(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

$\boldsymbol{\mathfrak{I}}$ が正則であるという条件のもとで，次式を得る。

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{\mathfrak{I}}^{-1}\boldsymbol{U}$$

$\boldsymbol{\mathfrak{I}}$ が定数とみなせるのであれば，$E[\boldsymbol{U}] = \boldsymbol{0}$ であることから，$E[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}] = \boldsymbol{0}$ となる。したがって，$\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の一致推定量である。

$\hat{\boldsymbol{\beta}}$ の分散共分散行列は，次式となる。

$$\begin{aligned}
E\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top\right] &= E\left[\boldsymbol{\mathfrak{I}}^{-1}\boldsymbol{U}(\boldsymbol{\mathfrak{I}}^{-1}\boldsymbol{U})^\top\right] \quad (\because \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{\mathfrak{I}}^{-1}\boldsymbol{U}) \\
&= E\left[\boldsymbol{\mathfrak{I}}^{-1}\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{\mathfrak{I}}^{-1}\right] \quad (\because \boldsymbol{\mathfrak{I}}^{-1} = (\boldsymbol{\mathfrak{I}}^{-1})^\top) \\
&= \boldsymbol{\mathfrak{I}}^{-1}E\left[\boldsymbol{U}\boldsymbol{U}^\top\right]\boldsymbol{\mathfrak{I}}^{-1} \\
&= \boldsymbol{\mathfrak{I}}^{-1} \quad (\because E\left[\boldsymbol{U}\boldsymbol{U}^\top\right] = V[\boldsymbol{U}] = \boldsymbol{\mathfrak{I}})
\end{aligned}$$

したがって，最尤推定量 $\hat{\boldsymbol{\beta}}$ は漸近的に多変量正規分布 $N_{p+1}(\boldsymbol{\beta}, \boldsymbol{\mathfrak{I}}^{-1})$ に従う。また，$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ は漸近的に多変量正規分布 $N_{p+1}(\boldsymbol{0}, \boldsymbol{\mathfrak{I}}^{-1})$ に従う。さらに，次式のワルド統計量を得る。

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top(\boldsymbol{\mathfrak{I}}^{-1}(\boldsymbol{\beta}))^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top\boldsymbol{\mathfrak{I}}(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi^2(p+1)$$

一般化線形モデルにおける応答変数が正規分布であるならば，$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ は正確に多変量正規分布 $N_{p+1}(\boldsymbol{0}, \boldsymbol{\mathfrak{I}}^{-1})$ に従う。さらに，ワルド統計量も正確に自由度 $p+1$ のカイ二乗分布 $\chi^2(p+1)$ に従う。

3.1 正規線形モデルの最尤推定量

$Y_1, \dots, Y_n$ は互いに独立な確率変数とし，次のモデルを考える。

$$E[Y_i] = \mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \quad Y_i \sim N(\mu_i, \sigma^2) \quad (i = 1, \dots, n)$$

このモデルは，連結関数として恒等関数 $g(\mu_i) = \mu_i$ をもつ一般化線形モデル（正規線形モデル）である。

連結関数は恒等関数であるので，$\mu_i = \eta_i$ であり，$\dfrac{\partial \mu_i}{\partial \eta_i} = 1$ $(i = 1 \dots, n)$ である。したがって，情報行列の要素は，次のようになる。

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

したがって，情報行列は次のようになる。

$$\boldsymbol{\mathfrak{I}} = \frac{1}{\sigma^2} \boldsymbol{X}^\top \boldsymbol{X}$$

同様に，

$$\begin{aligned} z_i &= \sum_{t=0}^{p} x_{ti} \hat{\beta}_t^{(m-1)} + (Y_i - \mu_i) \cdot \frac{\partial \eta_i}{\partial \mu_i} \\ &= \sum_{t=0}^{p} x_{ti} \hat{\beta}_t^{(m-1)} + (Y_i - \mu_i) \end{aligned}$$

となることから，$\hat{\boldsymbol{\beta}}^{(m-1)}$ に対して計算された μ_i は，$E[Y_i] = \mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$ より，$\boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}^{(m-1)} = \sum_{t=0}^{p} x_{ti} \hat{\beta}_t^{(m-1)}$ となる。したがって，$z_i = Y_i$ を得る。

以上の結果から，推定方程式 $\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X} \hat{\boldsymbol{\beta}}^{(m)} = \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{z}$ は，次のようになる。

$$\frac{1}{\sigma^2} \boldsymbol{X}^\top \boldsymbol{X} \hat{\boldsymbol{\beta}} = \frac{1}{\sigma^2} \boldsymbol{X}^\top \boldsymbol{Y}$$

したがって，$\boldsymbol{X}^\top \boldsymbol{X}$ が正則であれば，$\boldsymbol{\beta}$ の最尤推定量は次のようになる。

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$$

正規線形モデルは，次のように表現することができる。

$$\boldsymbol{Y} = \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

ここで，$\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$ とすると，$\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma \boldsymbol{I}_n)$ である。$\boldsymbol{\beta}$ の最尤推定量の期待値は次のよ

うになる。

$$
\begin{aligned}
E[\hat{\boldsymbol{\beta}}] &= E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}] \\
&= E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})] \\
&= E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}] + E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{\varepsilon}] \\
&= \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top E[\boldsymbol{\varepsilon}] \\
&= \boldsymbol{\beta}
\end{aligned}
$$

したがって，$\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の不偏推定量である。

また，最尤推定量 $\hat{\boldsymbol{\beta}}$ の分散 $V[\hat{\boldsymbol{\beta}}]$ は次式で与えられる。

$$
\begin{aligned}
V[\hat{\boldsymbol{\beta}}] &= V[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}] \\
&= V[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})] \\
&= V[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{\varepsilon}] \\
&= V[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{\varepsilon}] \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top V[\boldsymbol{\varepsilon}]((\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \sigma^2 \boldsymbol{I}_n \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \\
&= \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}
\end{aligned}
$$

したがって，$\sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \mathfrak{I}^{-1}$ となることから，$V[\hat{\boldsymbol{\beta}}] = \mathfrak{I}^{-1}$ であることが確認できる。

最尤推定量 $\hat{\boldsymbol{\beta}}$ は，$\boldsymbol{Y}$ の線形変換である。$\boldsymbol{Y} \sim N(\boldsymbol{0}, \sigma \boldsymbol{I}_n)$ であることから，$\hat{\boldsymbol{\beta}}$ の従う分布もまた正規分布となる。したがって，$\hat{\boldsymbol{\beta}}$ の正確な標本分布は，$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \mathfrak{I}^{-1})$ であり，ワルド統計量は $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathfrak{I}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi^2(p+1)$ である。

4 尤度比統計量

関心のあるモデルの適切さを評価する一つの方法は，推定されるパラメータの最大個数を含んだ**飽和モデル**（saturated model）とよばれる，最も一般的なモデルと比較することである。ただし，関心のあるモデルと飽和モデルは同じ確率分布および同じ連結関数をもつ一般化線形モデルである。

n 個の線形成分 $\boldsymbol{x}_i^\top \boldsymbol{\beta}$ に対して，n 個の観測値 y_i $(i = 1, \ldots, n)$ が得られているとする。このとき，飽和モデルは n 個のパラメータで表現される。飽和モデルは，**最大モデル**（maximal model）または**フルモデル**（full model）ともよばれる。

飽和モデルのパラメータ数を m とする。$\boldsymbol{\beta}_s$ を飽和モデルのパラメータベクトル，$\hat{\boldsymbol{\beta}}_s$ を $\boldsymbol{\beta}_s$ の最尤推定量とする。$\hat{\boldsymbol{\beta}}_s$ で評価した飽和モデルの尤度関数 $L(\hat{\boldsymbol{\beta}}_s; \boldsymbol{y})$ は，これらの観測値に対して同じ分布と同じ連結関数をもつ，他のどの尤度関数よりも大きくなる。その理由は，飽和モデルが最も完全にデータを記述しているからである。尤度関数 $L(\hat{\boldsymbol{\beta}}; \boldsymbol{y})$ を関心のあるモデルの尤度関数の最大値とする。このとき，モデルの適合度を評価する指標として，尤度比は次のように定義される。

$$
\lambda = \frac{L(\hat{\boldsymbol{\beta}}_s; \boldsymbol{y})}{L(\hat{\boldsymbol{\beta}}; \boldsymbol{y})}
$$

実際には，尤度比の対数，つまり，対数尤度の差が用いられる。

$$\log\lambda = \log\left(\frac{L(\hat{\boldsymbol{\beta}}_s;\boldsymbol{y})}{L(\hat{\boldsymbol{\beta}};\boldsymbol{y})}\right) = l(\hat{\boldsymbol{\beta}}_s;\boldsymbol{y}) - l(\hat{\boldsymbol{\beta}};\boldsymbol{y})$$

$\log\lambda$ が大きいとき，関心のあるモデルは飽和モデルと比較してよくデータを記述していないことを示唆する。また，$\log\lambda$ の棄却域を定めるためには標本分布が必要である。

次節にて，$2\log\lambda$ がカイ二乗分布に従うことを示す。したがって，$\log\lambda$ よりも $2\log\lambda$ の方が一般的に使われる統計量である。統計量 $2\log\lambda$ は，**逸脱度**（deviance）とよばれている。

5　逸脱度の標本分布

逸脱度は**対数尤度比統計量**（log likelihood ratio statistic）ともよばれ，次のように定義される。

$$D = 2\log\lambda = 2(l(\hat{\boldsymbol{\beta}}_s;\boldsymbol{y}) - l(\hat{\boldsymbol{\beta}};\boldsymbol{y}))$$

$\hat{\boldsymbol{\beta}}$ がパラメータ $\boldsymbol{\beta}$ の最尤推定量ならば，$\boldsymbol{U}(\hat{\boldsymbol{\beta}}) = \boldsymbol{0}$ であることから，次式が成り立つ。

$$\begin{aligned} l(\boldsymbol{\beta}) - l(\hat{\boldsymbol{\beta}}) &= -\frac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^\top \mathfrak{I}(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) \\ &\qquad \left(\because l(\boldsymbol{\beta}) = l(\hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^\top \boldsymbol{U}(\hat{\boldsymbol{\beta}}) - \frac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^\top \mathfrak{I}(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right) \end{aligned}$$

したがって，$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top \mathfrak{I}(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \sim \chi^2(p+1)$ であることから，

$$2(l(\hat{\boldsymbol{\beta}};\boldsymbol{y}) - l(\boldsymbol{\beta};\boldsymbol{y})) = (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^\top \mathfrak{I}(\boldsymbol{\beta})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})$$

は自由度 $p+1$ のカイ二乗分布 $\chi^2(p+1)$ に従う。ただし，$p+1$ は関心のあるモデルに含まれるパラメータ数である。

逸脱度は次のように表現できる。

$$\begin{aligned} D &= 2(l(\hat{\boldsymbol{\beta}}_s;\boldsymbol{y}) - l(\hat{\boldsymbol{\beta}};\boldsymbol{y})) \\ &= 2(l(\hat{\boldsymbol{\beta}}_s;\boldsymbol{y}) - l(\boldsymbol{\beta}_s;\boldsymbol{y})) - 2(l(\hat{\boldsymbol{\beta}};\boldsymbol{y}) - l(\boldsymbol{\beta};\boldsymbol{y})) + 2(l(\boldsymbol{\beta}_s;\boldsymbol{y}) - l(\boldsymbol{\beta};\boldsymbol{y})) \end{aligned}$$

第 1 項は，自由度 m のカイ二乗分布 $\chi^2(m)$ に従う。ここで，m は飽和モデルのパラメータ数である。第 2 項は，自由度 $p+1$ のカイ二乗分布 $\chi^2(p+1)$ に従う。第 3 項は，$\nu = 2(l(\boldsymbol{\beta}_s;\boldsymbol{y}) - l(\boldsymbol{\beta};\boldsymbol{y}))$ は関心のあるモデルが飽和モデルと同程度の適合度であるとき 0 に近い正の定数となる。したがって，逸脱度 D の標本分布は，近似的に自由度 $m-(p+1)$ の非心カイ二乗分布 $\chi^2(m-(p+1),\nu)$ に従う。ただし，ν は非心パラメータである。

応答変数 $Y_1,\ldots,Y_n$ が正規分布に従うとき，D は正確にカイ二乗分布に従う。しかし，D は未知である $V[Y_i]=\sigma^2$ に依存する。つまり，D は適合度を測る統計量として直接には使えないことを意味する。応答変数 $Y_1,\ldots,Y_n$ が他の分布に従う場合，D は近似的にカイ二乗分布に従う。しかし，二項分布やポアソン分布などの場合，D は実際に計算でき，適合度を測る統計量としてそのまま利用することができる。

5.1 正規線形モデルの逸脱度

$Y_1, \ldots, Y_n$ は互いに独立な確率変数とし，次のモデルを考える。

$$E[Y_i] = \mu_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \quad Y_i \sim N(\mu_i, \sigma^2) \quad (i = 1, \ldots, n)$$

このとき，対数尤度関数は次のようになる。

$$l(\boldsymbol{\beta}, \boldsymbol{y}) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu_i)^2 - \frac{1}{2} n \log(2\pi\sigma^2)$$

飽和モデルのもとでは，すべての μ_i が異なるので，$\boldsymbol{x}_i^\top \boldsymbol{\beta} = (\mu_1, \ldots, \mu_n)^\top$ となる。対数尤度関数を μ_i について偏微分し，推定方程式 $\dfrac{\partial l(\boldsymbol{\beta}, \boldsymbol{y})}{\partial \mu_i} = 0$ を解くと，飽和モデルのもとでの最尤推定値 $\hat{\mu}_i = y_i$ を得る。したがって，飽和モデルに対する対数尤度関数の最大値は次のようになる。

$$l(\hat{\boldsymbol{\beta}}_s, \boldsymbol{y}) = -\frac{1}{2} n \log(2\pi\sigma^2)$$

一方，パラメータ数 $p+1$ が $p < n$ となるモデルに対する最尤推定値は $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$ であるから，対応する対数尤度関数の最大値は次のようになる。

$$l(\hat{\boldsymbol{\beta}}, \boldsymbol{y}) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}})^2 - \frac{1}{2} n \log(2\pi\sigma^2)$$

したがって，逸脱度は次のようになる。

$$\begin{aligned} D &= 2(l(\hat{\boldsymbol{\beta}}_s; \boldsymbol{y}) - l(\hat{\boldsymbol{\beta}}; \boldsymbol{y})) \\ &= \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}})^2 \\ &= \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2 \end{aligned}$$

ここで，$\hat{\mu}_i = \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}}$ である。

逸脱度は次のようにも表すことができる。

$$\begin{aligned} D &= 2(l(\hat{\boldsymbol{\beta}}_s; \boldsymbol{y}) - l(\hat{\boldsymbol{\beta}}; \boldsymbol{y})) \\ &= \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}})^2 \\ &= \frac{1}{\sigma^2} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \end{aligned}$$

ここで，デザイン行列 $\boldsymbol{X}$ の行は $\boldsymbol{x}_i$ である。

$$\begin{aligned} \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} &= \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y} \\ &= (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top) \boldsymbol{Y} \\ &= (\boldsymbol{I}_n - \boldsymbol{H}) \boldsymbol{Y} \end{aligned}$$

ここで，$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ であり，$\boldsymbol{H}$ はハット行列とよばれる。ハット行列 $\boldsymbol{H}$ は射影行列であるので，$\boldsymbol{H} = \boldsymbol{H}^\top$ かつ $\boldsymbol{H}^2 = \boldsymbol{H}$ を満たす。したがって，次式が成り立つ。

$$
\begin{aligned}
(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) &= ((\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y})^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y} \\
&= \boldsymbol{Y}^\top(\boldsymbol{I}_n - \boldsymbol{H})^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y} \\
&= \boldsymbol{Y}^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{Y} \quad (\because \text{補足 2 を参照})
\end{aligned}
$$

補足 2

ハット行列 $\boldsymbol{H}$ が射影行列（対称行列かつべき等行列）であることを示す。まず，ハット行列 $\boldsymbol{H}$ が対称行列 $\boldsymbol{H} = \boldsymbol{H}^\top$ であることを示す。

$$
\begin{aligned}
\boldsymbol{H}^\top &= (\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top \\
&= \boldsymbol{X}((\boldsymbol{X}^\top\boldsymbol{X})^{-1})^\top\boldsymbol{X}^\top \\
&= \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top \quad (\because \boldsymbol{X}^\top\boldsymbol{X} = (\boldsymbol{X}^\top\boldsymbol{X})^\top \Rightarrow (\boldsymbol{X}^\top\boldsymbol{X})^{-1} = ((\boldsymbol{X}^\top\boldsymbol{X})^{-1})^\top) \\
&= \boldsymbol{H}
\end{aligned}
$$

次に，ハット行列 $\boldsymbol{H}$ がべき等行列 $\boldsymbol{H}^2 = \boldsymbol{H}$ であることを示す。

$$
\begin{aligned}
\boldsymbol{H}^2 &= \boldsymbol{H}\boldsymbol{H} \\
&= (\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top) \\
&= \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top \\
&= \boldsymbol{H}
\end{aligned}
$$

したがって，ハット行列 $\boldsymbol{H}$ が射影行列（対称行列かつべき等行列）である。
以上の結果を用いれば，簡単に $\boldsymbol{I}_n - \boldsymbol{H}$ も射影行列（対称行列かべき等行列）となることが示せる。

$$
\begin{aligned}
(\boldsymbol{I}_n - \boldsymbol{H})^\top &= \boldsymbol{I}_n^\top - \boldsymbol{H}^\top \\
&= \boldsymbol{I}_n - \boldsymbol{H}
\end{aligned}
$$

$$
\begin{aligned}
(\boldsymbol{I}_n - \boldsymbol{H})^2 &= \boldsymbol{I}_n^2 - \boldsymbol{I}_n\boldsymbol{H} - \boldsymbol{H}\boldsymbol{I}_n + \boldsymbol{H}^2 \\
&= \boldsymbol{I}_n - 2\boldsymbol{H} + \boldsymbol{H} \\
&= \boldsymbol{I}_n - \boldsymbol{H}
\end{aligned}
$$

　証明などの詳細は省略するが，一般に，$\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{V})$ であり，$\boldsymbol{V}$ が正則であるとき，確率変数 $\boldsymbol{Y}^\top\boldsymbol{V}^{-1}\boldsymbol{Y}$ は非心カイ二乗分布 $\chi^2(n, \nu)$ に従う。ただし，非心パラメータ $\nu = \boldsymbol{\mu}^\top\boldsymbol{V}^{-1}\boldsymbol{\mu}$ である。さらに，$\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{V})$ であり，$\boldsymbol{V}$ は $\mathrm{rank}(\boldsymbol{V}) = k < n$ で特異であるとき，$\boldsymbol{V}$ の一般化逆行列を $\boldsymbol{V}^-$ とすると，確率変数 $\boldsymbol{Y}^\top\boldsymbol{V}^-\boldsymbol{Y}$ は非心カイ二乗分布 $\chi^2(k, \nu)$ に従う。ただし，非心パラメータ $\nu = \boldsymbol{\mu}^\top\boldsymbol{V}^-\boldsymbol{\mu}$ である。証明などの詳細に関心がある場合は，Rao (1973) の第 3 章を参照すること。

　逸脱度 D の従う分布を考える。$\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma\boldsymbol{I}_n)$，$\boldsymbol{I}_n$ の階数は n，$\boldsymbol{H}$ の階数は $p+1$，$\boldsymbol{I}_n - \boldsymbol{H}$ の階数は $n-(p+1)$ であるから（補足 3 を参照），逸脱度 D は非心カイ二乗分布 $\chi^2(n-(p+1), \nu)$

に従う。ただし，非心パラメータ $\nu = \dfrac{(\boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{I}_n - \boldsymbol{H})(\boldsymbol{X}\boldsymbol{\beta})}{\sigma^2}$ である。しかしながら，

$$
\begin{aligned}
(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X} &= \boldsymbol{X} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X} \\
&= \boldsymbol{X} - \boldsymbol{X} \\
&= \boldsymbol{0}
\end{aligned}
$$

となることから，逸脱度 D は（中心）カイ二乗分布 $\chi^2(n-(p+1))$ に正確に従う。

補足 3

射影行列（対称行列かつべき等行列）の性質についてまとめる。
行列 $\boldsymbol{A}$ をべき等行列であるとする。べき等行列の固有値は 0 または 1 であることを示す。行列 $\boldsymbol{A}$ の固有値を λ とすると，固有値の定義から次式が成り立つ。

$$
\begin{aligned}
\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} &\Leftrightarrow \boldsymbol{A}^2\boldsymbol{x} = \boldsymbol{A}\lambda\boldsymbol{x} \\
&\Leftrightarrow \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{A}\boldsymbol{x} \quad (\because \boldsymbol{A}^2 = \boldsymbol{A}) \\
&\Leftrightarrow \lambda\boldsymbol{x} = \lambda^2\boldsymbol{x} \quad (\because \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}) \\
&\Leftrightarrow \lambda(\lambda-1)\boldsymbol{x} = \boldsymbol{0}
\end{aligned}
$$

$\boldsymbol{x} \neq \boldsymbol{0}$ であることから，$\lambda = 0, 1$ を得る。つまり，べき等行列の固有値は 0 または 1 である。
射影行列 $\boldsymbol{A}$ について，$\mathrm{tr}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A})$ が成り立つことを示す。行列 $\boldsymbol{A}$ は対称行列であるので，適当な直交行列 $\boldsymbol{P}$ を用いて，次のように対角化できる。

$$
\boldsymbol{P}^\top\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D}
$$

ここで，$\boldsymbol{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ であり，$\lambda_1, \ldots, \lambda_n$ は行列 $\boldsymbol{A}$ の固有値である。行列 $\boldsymbol{P}^\top\boldsymbol{A}\boldsymbol{P}$ トレースは，次のようになる。

$$
\mathrm{tr}(\boldsymbol{P}^\top\boldsymbol{A}\boldsymbol{P}) = \mathrm{tr}(\boldsymbol{D}) = \sum_{i=1}^{n}\lambda_i
$$

一方，トレースの性質と $\boldsymbol{P}$ が直交行列であることから，次式を得る。

$$
\mathrm{tr}(\boldsymbol{P}^\top\boldsymbol{A}\boldsymbol{P}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{P}\boldsymbol{P}^\top) = \mathrm{tr}(\boldsymbol{A})
$$

したがって，対称行列であれば，$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{n}\lambda_i$ が成り立つ。さらに，べき等行列の固有値は 0 または 1 であることから，射影行列 $\boldsymbol{A}$ の $\mathrm{tr}(\boldsymbol{A})$ は固有値 1 の数に等しいことがわかる。一方，$\boldsymbol{D}$ の階数について考えると，行列の基本変形をすれば，$\mathrm{rank}(\boldsymbol{D}) = \mathrm{rank}(\mathrm{diag}(1, \cdots 1, 0 \ldots, 0))$ となる。つまり，$\mathrm{rank}(\boldsymbol{D})$ は，射影行列 $\boldsymbol{A}$ の固有値 1 の数に等しいことがわかる。

$$
\mathrm{rank}(\boldsymbol{D}) = \mathrm{rank}(\boldsymbol{P}^\top\boldsymbol{A}\boldsymbol{P}) = \mathrm{rank}(\boldsymbol{A})
$$

であることから，射影行列 $\boldsymbol{A}$ について，$\mathrm{tr}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A})$ は成り立つ。

6 仮説検定

長さ $p+1$ のパラメータベクトル $\boldsymbol{\beta}=(\beta_0,\beta_1,\dots,\beta_p)^\top$ に関する仮説は，ワルド統計量

$$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\mathfrak{I}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\sim\chi^2(p+1)$$

を用いて検定できる。スコア統計量 $\boldsymbol{U}^\top\mathfrak{I}^{-1}\boldsymbol{U}\sim\chi^2(p+1)$ が使用されることもある。

別のアプローチとして，2 つのモデルの適合度を比較する方法がある。この 2 つのモデルは**入れ子**（nested）あるいは**階層的**（hierarchical）になっている必要がある。つまり，同じ確率分布および同じ連結関数をもち，単純な方のモデル M_0 の線形成分が，より一般的なモデル M_1 の線形成分の特別な場合となっている必要がある。

モデル M_0 に対応する帰無仮説 H_0 と，モデル M_1 に対応するより一般的な仮説（対立仮説）H_1 を次のように考える。

$$H_0:\boldsymbol{\beta}=\boldsymbol{\beta}_0=\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_q\end{pmatrix},\quad H_1:\boldsymbol{\beta}=\boldsymbol{\beta}_1=\begin{pmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_p\end{pmatrix}$$

ただし，$q<p<n$ である。

対立仮説 H_1 に対する帰無仮説 H_0 の検定は，次のような逸脱度の差 ΔD を用いることができる。

$$\begin{aligned}\Delta D&=D_0-D_1\\&=2(l(\hat{\boldsymbol{\beta}}_s;\boldsymbol{y})-l(\hat{\boldsymbol{\beta}}_0;\boldsymbol{y}))-2(l(\hat{\boldsymbol{\beta}}_s;\boldsymbol{y})-l(\hat{\boldsymbol{\beta}}_1;\boldsymbol{y}))\\&=2(l(\hat{\boldsymbol{\beta}}_1;\boldsymbol{y})-l(\hat{\boldsymbol{\beta}}_0;\boldsymbol{y}))\end{aligned}$$

M_0 と M_1 のどちらもデータをよく記述しているならば，$D_0\sim\chi^2(n-(q+1))$ および $D_1\sim\chi^2(n-(p+1))$ が成り立つ。したがって，$\Delta D\sim\chi^2(p-q)$ となる（補足 4 を参照）。観測値から計算された ΔD の値が $\chi^2(p-q)$ に矛盾しないならば，より単純であるという理由で H_0 に対応するモデル M_0 を選択する。ΔD の値が棄却域に入っていれば，モデル M_1 の方がモデル M_0 よりもデータを有意によく記述している（M_1 がデータに特に適法していない場合でも）として，H_1 を選び H_0 を棄却する。

逸脱度がデータから計算できる場合には，逸脱度の差 ΔD は仮説検定の便利な方法を与える。ΔD の標本分布は，通常，逸脱度そのものの標本分布よりもカイ二乗分布の近似が良いとされている。正規分布に基づくモデルや，推定されない局外パラメータをもつ他の分布の場合，逸脱度はデータから完全に決まらないかもしれない。次に，正規線形モデルの場合を考えることで，この問題の解決法を例示する。

補足 4

Cochran の定理より，次の定理が成り立つ。
2 つの確率変数 $X_1^2 \sim \chi^2(m)$ と $X_2^2 \sim \chi^2(k)$ を考える。ただし，$m > k$ とする。$X^2 = X_1^2 - X_2^2$ が非負定値であるならば，$X^2 \sim \chi^2(m-k)$ である。

M_0 と M_1 のどちらもデータをよく記述しているならば，$D_0 \sim \chi^2(n-(q+1))$ および $D_1 \sim \chi^2(n-(p+1))$ が成り立つ。このとき，$\Delta D \sim \chi^2(p-q)$ となることを示す。ここで，$q < p < n$ より，$n-(q+1) > n-(p+1)$ である。

$$\begin{aligned}\Delta D &= D_0 - D_1 \\ &= \frac{1}{\sigma^2}\left(\boldsymbol{Y}^\top(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{Y} - \boldsymbol{Y}^\top(\boldsymbol{I}_n - \boldsymbol{H}_1)\boldsymbol{Y}\right)\end{aligned}$$

ここで， $\boldsymbol{H}_0 = \boldsymbol{X}_0(\boldsymbol{X}_0^\top\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0^\top$ と $\boldsymbol{H}_1 = \boldsymbol{X}_1(\boldsymbol{X}_1^\top\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top$ であり，

$$\boldsymbol{X}_0 = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{q1} \\ 1 & x_{12} & x_{22} & \dots & x_{q2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{qn} \end{pmatrix}, \quad \boldsymbol{X}_1 = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{pmatrix}$$

である。$\boldsymbol{Y}^\top(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{Y}$ と $\boldsymbol{Y}^\top(\boldsymbol{I}_n - \boldsymbol{H}_1)\boldsymbol{Y}$ はスカラーであり，$q < p < n$ より，$\boldsymbol{Y}^\top(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{Y} - \boldsymbol{Y}^\top(\boldsymbol{I}_n - \boldsymbol{H}_1)\boldsymbol{Y} \geq 0$ となる。したがって，$\boldsymbol{Y}^\top(\boldsymbol{I}_n - \boldsymbol{H}_0)\boldsymbol{Y} - \boldsymbol{Y}^\top(\boldsymbol{I}_n - \boldsymbol{H}_1)\boldsymbol{Y}$ は非負定値行列である。以上の結果から，$\Delta D \sim \chi^2(p-q)$ となる。

6.1 正規線形モデルの仮説検定

$Y_1, \dots, Y_n$ は互いに独立な確率変数が次の正規線形モデルに従うとする。

$$E[Y_i] = \mu_i = \boldsymbol{x}_i^\top\boldsymbol{\beta}, \quad Y_i \sim N(\mu_i, \sigma^2) \quad (i = 1, \dots, n)$$

このとき，逸脱度は次のようになる。

$$\begin{aligned}D &= 2(l(\hat{\boldsymbol{\beta}}_s; \boldsymbol{y}) - l(\hat{\boldsymbol{\beta}}; \boldsymbol{y})) \\ &= \frac{1}{\sigma^2}\sum_{i=1}^n (y_i - \boldsymbol{x}_i^\top\hat{\boldsymbol{\beta}})^2 \\ &= \frac{1}{\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})\end{aligned}$$

$\hat{\boldsymbol{\beta}}_0$ と $\hat{\boldsymbol{\beta}}_1$ をそれぞれ，帰無仮説 H_0 に対応するモデル M_0 と対立仮説 H_1 に対応するモデル M_1 のパラメータベクトルの最尤推定量とする。$\boldsymbol{X}_0$ と $\boldsymbol{X}_1$ をそれぞれ，モデル M_0 と M_1 に対応するデザイン行列とする。

このとき，モデル M_0 と M_1 の逸脱度 D_0 と D_1 はそれぞれ次のようになる。

$$
\begin{aligned}
D_0 &= \frac{1}{\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}_0\hat{\boldsymbol{\beta}}_0)^\top(\boldsymbol{Y} - \boldsymbol{X}_0\hat{\boldsymbol{\beta}}_0) \\
&= \frac{1}{\sigma^2}\left(\boldsymbol{Y}^\top\boldsymbol{Y} - \boldsymbol{Y}^\top\boldsymbol{X}_0\hat{\boldsymbol{\beta}}_0 - \hat{\boldsymbol{\beta}}_0^\top\boldsymbol{X}_0^\top\boldsymbol{Y} + \hat{\boldsymbol{\beta}}_0^\top\boldsymbol{X}_0^\top\boldsymbol{X}_0\hat{\boldsymbol{\beta}}_0\right) \\
&= \frac{1}{\sigma^2}\left(\boldsymbol{Y}^\top\boldsymbol{Y} - 2\hat{\boldsymbol{\beta}}_0^\top\boldsymbol{X}_0^\top\boldsymbol{Y} + \hat{\boldsymbol{\beta}}_0^\top\boldsymbol{X}_0^\top\boldsymbol{X}_0\hat{\boldsymbol{\beta}}_0\right) \\
&= \frac{1}{\sigma^2}\left(\boldsymbol{Y}^\top\boldsymbol{Y} - \hat{\boldsymbol{\beta}}_0^\top\boldsymbol{X}_0^\top\boldsymbol{Y}\right) \quad (\because \boldsymbol{X}_0^\top\boldsymbol{X}_0\hat{\boldsymbol{\beta}}_0 = \boldsymbol{X}_0^\top\boldsymbol{Y})
\end{aligned}
$$

$$
\begin{aligned}
D_1 &= \frac{1}{\sigma^2}(\boldsymbol{Y} - \boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1)^\top(\boldsymbol{Y} - \boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1) \\
&= \frac{1}{\sigma^2}\left(\boldsymbol{Y}^\top\boldsymbol{Y} - \hat{\boldsymbol{\beta}}_1^\top\boldsymbol{X}_1^\top\boldsymbol{Y}\right) \quad (\because \boldsymbol{X}_1^\top\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 = \boldsymbol{X}_1^\top\boldsymbol{Y})
\end{aligned}
$$

逸脱度の差 ΔD は次のようになる。

$$
\begin{aligned}
\Delta D &= D_0 - D_1 \\
&= \frac{1}{\sigma^2}\left[\left(\boldsymbol{Y}^\top\boldsymbol{Y} - \hat{\boldsymbol{\beta}}_0^\top\boldsymbol{X}_0^\top\boldsymbol{Y}\right) - \left(\boldsymbol{Y}^\top\boldsymbol{Y} - \hat{\boldsymbol{\beta}}_1^\top\boldsymbol{X}_1^\top\boldsymbol{Y}\right)\right] \\
&= \frac{1}{\sigma^2}\left(\hat{\boldsymbol{\beta}}_1^\top\boldsymbol{X}_1^\top\boldsymbol{Y} - \hat{\boldsymbol{\beta}}_0^\top\boldsymbol{X}_0^\top\boldsymbol{Y}\right)
\end{aligned}
$$

一般的に，M_1 はデータに良く適合する（対立仮説 H_1 は正しい）と仮定される。つまり，$D_1 \sim \chi^2(n-(p+1))$ であると仮定する。もし，M_0 もデータによく適合すれば，$\Delta D \sim \chi^2(p-q)$ となる。一方，M_0 がデータに適合しないならば，ΔD は非心カイ二乗分布に従う。ここで，σ^2 は未知であるため，逸脱度をデータから計算することができないことに注意する。この問題を解決するために，次式を用いる。

$$
F = \frac{\dfrac{D_0 - D_1}{p-q}}{\dfrac{D_1}{n-(p+1)}}
$$

ここで，F は当てはめ値から直接計算できることに注意する。H_0 が正しいならば，F は中心 $F(p-q, n-(p+1))$ 分布に従う。一方，H_0 が正しくないならば，F は中心 $F(p-q, n-(p+1))$ 分布から期待される値よりも大きくなる。

ここで述べられた仮説検定は，表 1 に示される分散分析表として要約されることが多い。

表 1　分散分析表

変動要因	自由度	平方和	平均平方和
$\boldsymbol{\beta}_0$ のモデル	q	$\hat{\boldsymbol{\beta}}_0^\top\boldsymbol{X}_0^\top\boldsymbol{Y}$	
$\boldsymbol{\beta}_1$ のモデルによる改善	$p-q$	$\hat{\boldsymbol{\beta}}_1^\top\boldsymbol{X}_1^\top\boldsymbol{Y} - \hat{\boldsymbol{\beta}}_0^\top\boldsymbol{X}_0^\top\boldsymbol{Y}$	$\dfrac{\hat{\boldsymbol{\beta}}_1^\top\boldsymbol{X}_1^\top\boldsymbol{Y} - \hat{\boldsymbol{\beta}}_0^\top\boldsymbol{X}_0^\top\boldsymbol{Y}}{p-q}$
残差	$n-p$	$\boldsymbol{Y}^\top\boldsymbol{Y} - \hat{\boldsymbol{\beta}}_1^\top\boldsymbol{X}_0^\top\boldsymbol{Y}$	$\dfrac{\boldsymbol{Y}^\top\boldsymbol{Y} - \hat{\boldsymbol{\beta}}_1^\top\boldsymbol{X}_0^\top\boldsymbol{Y}}{n-q}$
計	n	$\boldsymbol{Y}^\top\boldsymbol{Y}$	

7 適用例

表 2 は，高炭水化物ダイエットを 6 ヶ月間続けている 20 名の男性インスリン依存性糖尿病患者が摂取した複合炭水化物の総カロリーに占める割合を示している。

表 2　20 名の男性インスリン依存性糖尿病患者の炭水化物摂取量，年齢，相対体重，および蛋白摂取量

炭水化物 $\boldsymbol{y}$	年齢 $\boldsymbol{x}_1$	体重 $\boldsymbol{x}_2$	蛋白 $\boldsymbol{x}_3$
33	33	100	14
40	47	92	15
37	49	135	18
27	35	144	12
30	46	140	15
43	52	101	15
34	62	95	14
48	23	101	17
30	32	98	15
38	42	105	14
50	31	108	17
51	61	85	19
30	63	130	19
36	40	127	20
41	50	109	15
42	64	107	16
46	56	117	18
24	61	100	13
35	48	118	18
37	28	102	14

このダイエット療法に対するコンプライアンス（遵守の度合い）は，年齢（歳），体重（身長から計算される理想体重に対する割合），および摂取蛋白の総カロリーに占める割合といった他の要因に関係があると考えられている。そこで，これらの変数と説明変数とする次の正規線形モデル（線形回帰モデル）をデータに当てることにする。

$$E[Y_i] = \mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}, \quad Y_i \sim N(\mu_i, \sigma^2) \tag{1}$$

このモデルでは，年齢 $\boldsymbol{x}_1$，体重 $\boldsymbol{x}_2$，蛋白 $\boldsymbol{x}_3$ は炭水化物 $\boldsymbol{y}$ に線形に関係していると仮定している。

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_3 \end{pmatrix}$$

として，データから計算すると $\boldsymbol{X}^\top \boldsymbol{y}$ と $\boldsymbol{X}^\top \boldsymbol{X}$ は次のようになる。

$$\boldsymbol{X}^\top \boldsymbol{y} = \begin{pmatrix} 752 \\ 34596 \\ 82270 \\ 12105 \end{pmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{X} = \begin{pmatrix} 20 & 923 & 2214 & 318 \\ 923 & 45697 & 102003 & 14780 \\ 2214 & 102003 & 250346 & 35306 \\ 318 & 14780 & 35306 & 5150 \end{pmatrix}$$

したがって，$\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{y}$ の解は次のようになる。

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 36.9601 \\ -0.1137 \\ -0.2280 \\ 1.9577 \end{pmatrix}$$

逸脱度の使い方を説明するために，炭水化物 $\boldsymbol{y}$ は年齢 $\boldsymbol{x}_1$ には依存しないとして，$H_0 : \beta_1 = 0$ という仮説を検定する。対応するモデルは次のようになる。

$$E[Y_i] = \mu_i = \beta_0 + \beta_2 x_{2i} + \beta_3 x_{3i}, \quad Y_i \sim N(\mu_i, \sigma^2) \tag{2}$$

(2) 式に対応するデザイン行列 $\boldsymbol{X}$ を得るためには，(1) 式のモデルのデザイン行列から 2 列目を除けばよいので，$\boldsymbol{X}^\top \boldsymbol{y}$ と $\boldsymbol{X}^\top \boldsymbol{X}$ は次のようになる。

$$\boldsymbol{X}^\top \boldsymbol{y} = \begin{pmatrix} 752 \\ 82270 \\ 12105 \end{pmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{X} = \begin{pmatrix} 20 & 2214 & 318 \\ 2214 & 250346 & 35306 \\ 318 & 35306 & 5150 \end{pmatrix}$$

したがって，$\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{y}$ の解は次のようになる。

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 33.130 \\ -0.222 \\ 1.824 \end{pmatrix}$$

仮説 H_0 に対する有意性検定の結果は，表 3 に要約されている。検定統計量は $F = 38.36/35.48 = 1.08$ となり，これは $F(1, 16)$ の分布と比べて有意ではない。したがって，仮説 H_0 は棄却されず，炭水化物 $\boldsymbol{y}$ は年齢 $\boldsymbol{x}_1$ には依存しないと結論付けられる。

表 3　モデル (1) とモデル (2) を比較する分散分析表

変動要因	自由度	平方和	平均平方和
モデル (2)	3	28761.978	
モデル (1) による改善	1	38.359	38.36
残差	16	567.663	35.48
計	20	29368.000	

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.

[3] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications, 2nd Edition.* Wiley, New York.