2017年6月15日(金)-16日(土)　理論合宿@日本橋

統計モデルと一般化推定方程式

推定量の一致性と漸近正規性の証明

・Wedderburn, R. W., Quasi-likelihood functions, generalized linear models and the Gauss－Newton method. Biometrika, Vol.61, No.3, pp.439-447, 1974.
・Liang, K. Y., Zeger, S. L., Longitudinal data analysis using generalized linear models. Biometrika, Vol.73, No.1, pp.13-22, 1986.

2017年9月16日(金) 14:00～14:40

M1 井澤 佳那子

本日の発表内容

1. 最近の私の研究について
   ：選択行動における量子観測の情報系とその理論化

2. 統計モデルと推定

3. 一般化線形モデルと最尤推定

4. 擬似尤度スコア方程式（Wedderburn, 1974）

5. 一般化推定方程式（Zeger, S. L., Liang, K. Y. , 1986）

モチベーション

都市内の**<u>交通行動</u>**や**<u>都市の形成</u>**に対して**自然な表現を与える**ことで評価・分析を行い，それに基づいた都市交通計画立案に繋げたい.

特に交通行動について，**観測量の代数的構造**に自然な表現を与えることが，現象（都市内の交通行動原理）へのより良い理解に繋がる.

cf. 最大エントロピー原理

背景：観測と行動理論

1. 観測

・**データの多様化（quantization, sampling）**

：情報通信・観測技術の進展やセンサネットワークの普及

・**情報量の増大**

：**点単位での**行動把握が可能となり，交通行動に関する情報量が増大

・**点データの特徴量**

：点データの特徴量に合わせた空間の集計単位とスケールに焦点を当てた整理（+それら正規化する行動判別手法）が必要.

・**情報量的解釈の必要性**

：点ベースの交通行動データについて，**情報量を定義**し評価することで，観測量によるモデル性能を直感的に理解することができ，他の観測データとの比較や異なるモビリティ間の比較が容易になる.

背景：観測と行動理論

2. 行動理論

- **任意の選択行動に対する評価尺度**

  ：既出の行動モデルは，実質的に特定のモビリティを対象としているものや実際に真のパラメータを推定することが困難なもの.

- **情報量の異質性と相関**

  ：パラメータ推定では**観測データから得られる情報量が十分に大きいことが前提**.　交通行動の相関性により当該情報量が減少する可能性.

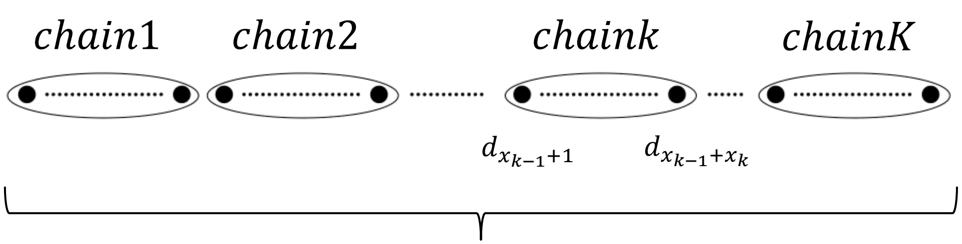
- **推定上の特性，計算性向上**

  ：交通行動判別においても行動モデルのパラメータ推定においても，情報量を定義することによる計算性の向上に対する寄与.

交通行動分析における量子観測の情報系を再整理.

フレームの検討

選択行動の情報量を**<u>選択結果チェーンが選択行動に関して持つ情報量と定義</u>**し，選択モデルのパラメータの全体分布とチェーンから推定されるパラメータの部分分布の相対性を利用して，当該**<u>情報量を実験的に評価</u>**する．



観測遷移数 $L = \sum_i^K x_i$

- 任意の空間単位に紐づけられる選択結果の1遷移

e.g. 交通分析の情報系アプローチ
(Sopasakis, A. et al., 2016)

1. 序論
2. 行動モデルの定式化
3. 情報量的解釈
4. 情報量規準（モデル選択）
5. 推定，パラメータの特性
6. 実データ推定
7. 結論

情報理論（Shannon, 1948）[基礎]

情報：不確実性を減少させてくれるもの.

性質① 発生確率の低いことを知れた方が情報量が多い.
性質② 情報量は足し合わされる.

- **エントロピー**

$$H(y) \coloneqq -\int q(y) \log q(y)\, dy$$

- **Kullback-Leibler ダイバージェンス**

$$D[q(y) : p(y)] \coloneqq \int q(y) \log \frac{q(y)}{p(y)}\, dy$$
$$= \int q(y) \log q(y)\, dy - \int q(y) \log p(y)\, dy$$

$p(y), q(y)$：確率変数yの確率分布

選択行動における情報量的解釈

今，$q(y)$が真の確率分布，$p(y;\beta)$が推定した確率分布とすると

$$D[q(y):p(y;\beta)] = \int q(y)\log q(y)\,dy - \int q(y)\log p(y;\beta)\,dy$$

一定

$\int q(y)\log p(y;\beta)\,dy$ が大きいと$D[q(y):p(y;\beta)]$は小さくなる．

$$\frac{1}{K}\int \log p(y;\beta)\,dy \rightarrow \int q(y)\log p(y;\beta)\,dy \quad (\text{データ数 } K \rightarrow \infty)$$

$\int \log p(y;\beta)\,dy$ をパラメタβについて最大化すれば良い[最尤推定法].

- 選択行動(e.g. MNL)では**これが選択確率**になっている．
- 配分(e.g. Dial, 1971)は，ローディングの問題なので単に**ネットワーク上のエントロピー最小化問題**.

情報量的解釈の意義

- 大数の法則の仮定の妥当性.

$$\frac{\mathbf{1}}{\boldsymbol{K}}\int \boldsymbol{log}\, \boldsymbol{p}(\boldsymbol{y};\boldsymbol{\beta})\, \boldsymbol{dy} \to \int \boldsymbol{q}(\boldsymbol{y})\, \mathbf{log}\, \boldsymbol{p}(\boldsymbol{y};\boldsymbol{\beta})\, \boldsymbol{dy} \quad (\text{データ数 } \boldsymbol{K} \to \infty)$$

- 本来最小化すべきは，KLダイバージェンス.

$$D[q(y){:}\,p(y;\beta)] = \int q(y)\log q(y)\,dy - \int q(y)\log p(y;\beta)\,dy$$

観測量の多様化に伴う
真の分布に対する仮定

モデル発展に伴う
母数の高次元化

- 任意の選択行動に対する評価尺度の一般化.

・Wedderburn, R. W., Quasi-likelihood functions, generalized linear models and the Gauss—Newton method. Biometrika, Vol.61, No.3, pp.439-447, 1974.
・Liang, K. Y., Zeger, S. L., Longitudinal data analysis using generalized linear models. Biometrika, Vol.73, No.1, pp.13-22, 1986.

統計モデルと推定

推定とは確率変数$\boldsymbol{y}$がある分布族$\boldsymbol{p(y,\beta)}$に従うとして，観測したデータ$\boldsymbol{y}$をもとに分布を求めること．

線形回帰モデル（最小線形二乗法）

↓

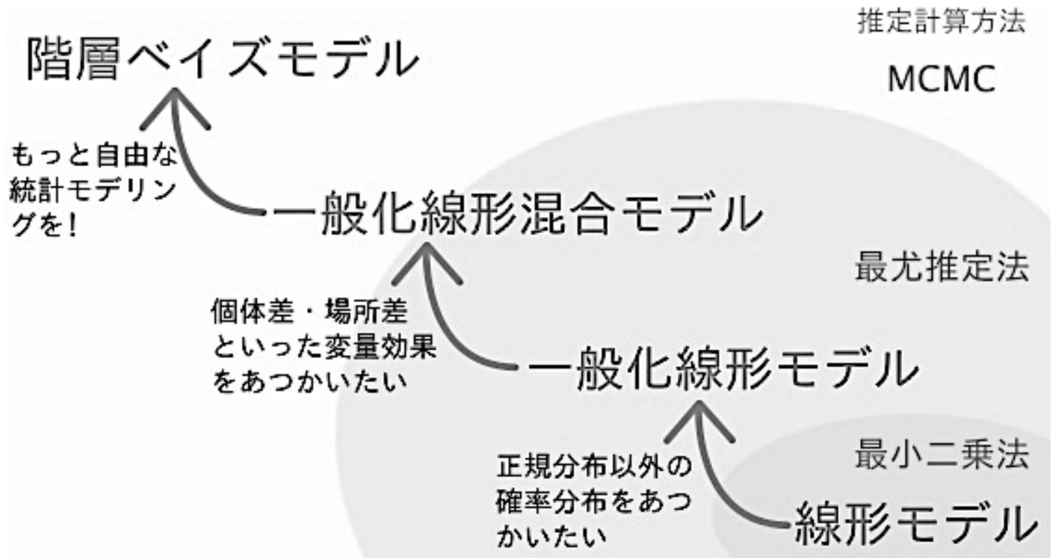
一般化線形モデル（最尤推定法）

↓ 母数の次元高度化

GEE，GMM（Newton法など）

↓

ベイズモデル（MCMCなど）



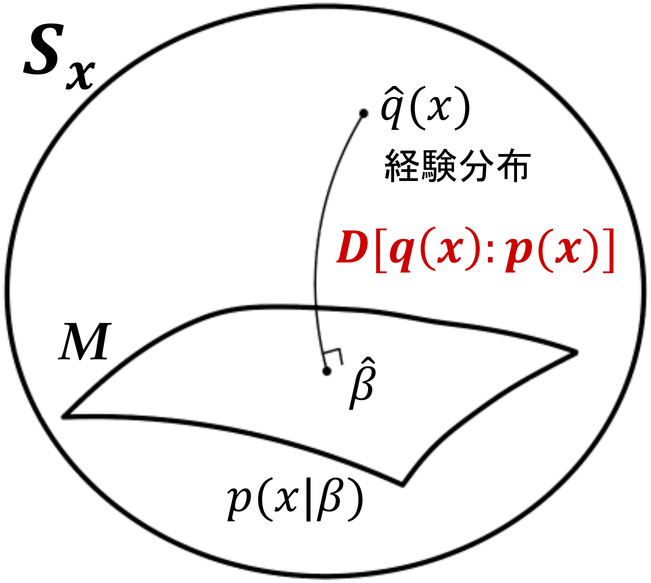
久保拓弥. (2012). *データ解析のための統計モデリング入門* (pp. 106–108). 岩波書店.

推定と幾何学的理解

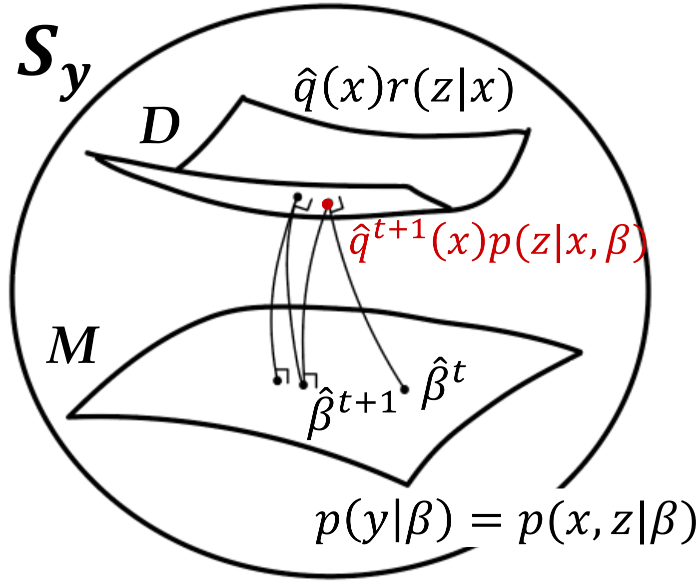
確率変数の一部は観測されないことが多い….

確率変数は，$\boldsymbol{y} = (\boldsymbol{x}, \boldsymbol{z})$のよう分割され，$\boldsymbol{x}$は観測できて$\boldsymbol{z}$は観測できない隠れ変数であるとする．（$\boldsymbol{S}$：確率分布空間）

隠れ変数のない場合 e.g. **GEE**



隠れ変数のある場合 e.g. **em**

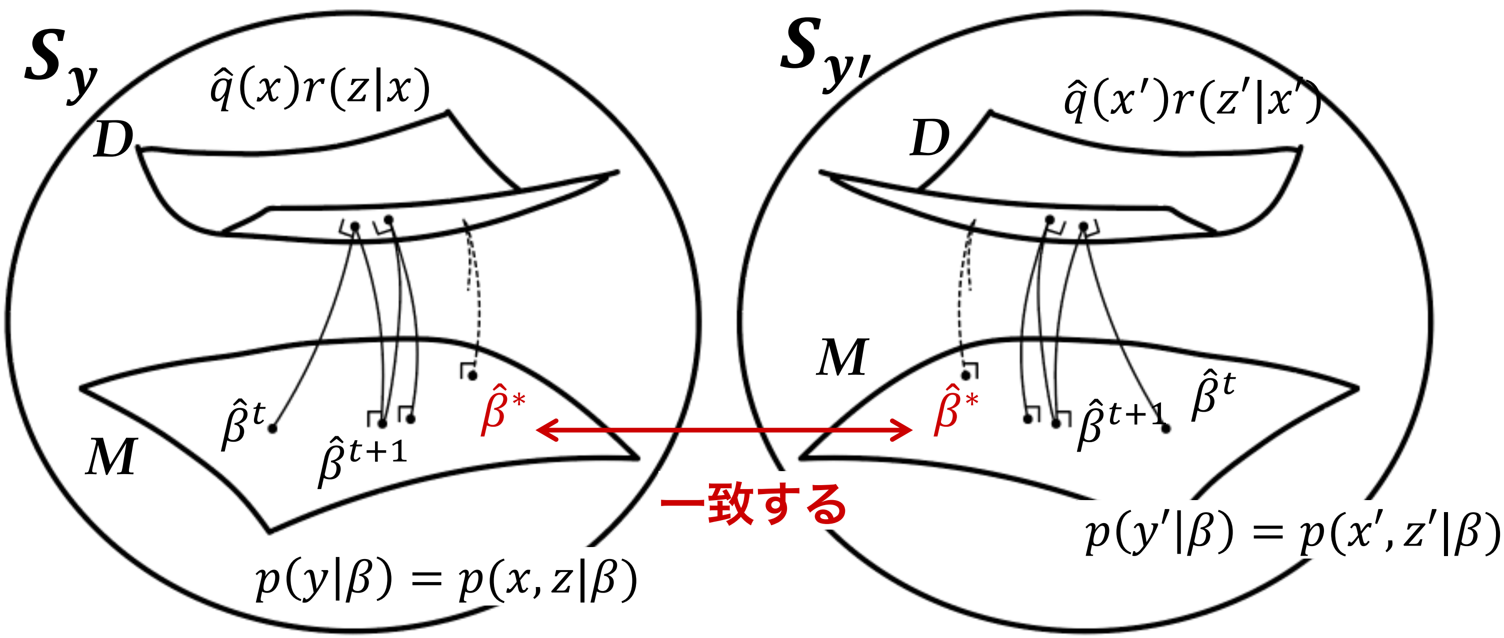


推定と幾何学的理解【おまけ】

確率変数の一部は観測されないことが多い….

確率変数は，$\boldsymbol{y} = (\boldsymbol{x}, \boldsymbol{z})$のよう分割され，$\boldsymbol{x}$は観測できて$\boldsymbol{z}$は観測できない隠れ変数であるとする．（$\boldsymbol{S}$：確率分布空間）

構造推定 理論モデルのパラメータを推定する



一般化推定方程式の背景

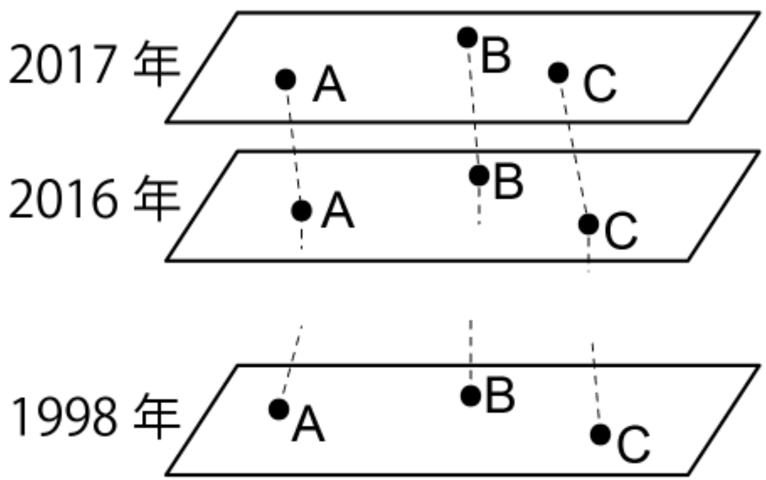
数学的には，一般化線形モデル(GLM)の拡張である**<u>擬似尤度スコア推定方程式</u>**の特殊な場合.

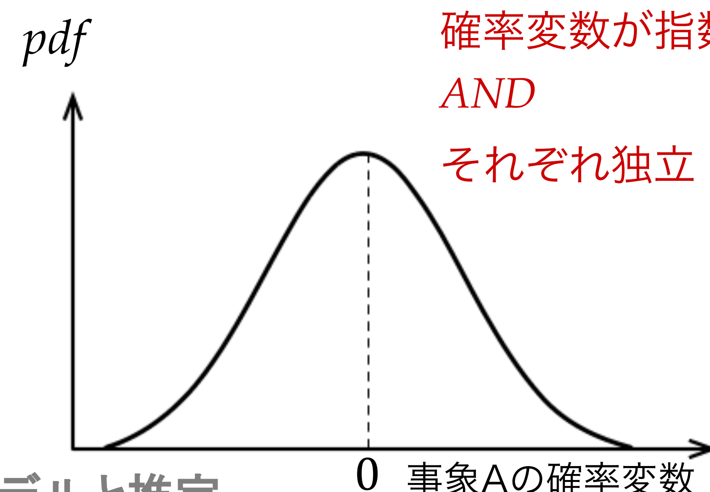
GLMの仮定

<u>仮定①</u>：従属変数が指数分布族に従う.

<u>仮定②</u>：従属変数がそれぞれ独立である.　尤度関数 $L(\theta|y) = \prod_{i=1}^{K} f(y_i)$

e.g. 継時的，繰り返し測定，クラスタデータ





確率変数が指数分布に従う
AND
それぞれ独立

表記の整理

- $\boldsymbol{y_{it}}$：従属変数
  $i = 1, \dots, K$：対象の数
  $t = 1, \dots, n_i$：観測回数
- $\boldsymbol{x_{it}}$：p×1の共分散ベクトル
- $\boldsymbol{\mu_i = E(y_{it}|x_{it})}$：期待値
- $\boldsymbol{\beta}$： p×1の回帰係数ベクトル
- $\boldsymbol{Y_i = \left(y_{i1}, \dots, y_{in_i}\right)^T}$：$n_i$×1の従属変数ベクトル
- $\boldsymbol{X_i = \left(x_{i1}, \dots, x_{in_i}\right)^T}$：$n_i$×$p$の$i(i = 1, \dots, K)$番目の共分散ベクトル
- $\boldsymbol{\theta_{it} = h(\eta_{it})}$：パラメータ
- $\boldsymbol{\eta_{it} = x_{it}\beta}$

GLMと最尤推定

Step1. 確率変数$\boldsymbol{Y_i}$の確率分布を仮定.

指数分布族は一般に,

$$f(y_i) = \exp\{\{y_i\theta_i - a(\theta_i) + b(y_i)\}\phi\}$$

Step2. 確率変数$\boldsymbol{Y_i}$の期待値$\boldsymbol{\mu_i}$の回帰モデルを設定.

$$\mu_i = E(y_{it}|x_{it}) = a'(\theta_{it}) = a'\left(h(\eta_{it})\right) = a'(h(x_{it}\beta))$$

Step3. 仮定した確率分布から尤度関数$\mathbf{L}(\boldsymbol{\theta}|\boldsymbol{y})$を作り，これを最大化する$\boldsymbol{\beta}(\widehat{\boldsymbol{\beta}})$を求める.

→対数尤度関数をβについて偏微分して極値を求める. e.g. $\frac{\partial}{\partial\boldsymbol{\beta}}\log\mathrm{L}(\theta|y) = 0$

スコア関数

$$\boldsymbol{U_I(\beta)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{K}} \boldsymbol{X_i}^{\boldsymbol{T}}\Delta_i\boldsymbol{S_i} = \mathbf{0}$$

$$where \qquad \Delta_i = diag(d\theta_{it}/d\eta_{it}),\ S_i = Y_i - a'(\theta_{it})$$

最尤推定量$\widehat{\beta}$の性質

<u>性質①.</u> 一致性

$$\text{確率1で}, \hat{\beta}_K \to \beta^* \quad (K \to \infty)$$

<u>性質②.</u> 漸近正規性

$$\sqrt{K}\left(\hat{\beta}_K - \beta^*\right) \to N(0, I(\beta^*)^{-1}) \quad (K \to \infty)$$

$\hat{\beta}_K$：推定値, β^* ：真値, K：対象数

擬似尤度スコア推定方程式の着想

(Wedderburn, 1974)

従属変数の真の分布が，仮定した分布でない場合にも**妥当な**統計的推論を行いたい.

e.g. ポアソン回帰モデルのover-dispersion問題

GLMの仮定の再確認

仮定①：従属変数が指数分布族に従う. **←この仮定①を緩める**

仮定②：従属変数がそれぞれ独立である. 尤度関数 $L(\theta|y) = \prod_{i=1}^{K} f(y_i)$

Y_iの従う確率分布が指数分布族であるとき平均と分散は,

$$\boldsymbol{E(y_i) = \mu_i = a'(\theta_i)},\quad \boldsymbol{Var(Y_i) = a''(\theta_i)/\emptyset}$$

で，任意の関数 $v(\cdot)$ によって関係は一意に決まる.

$$\boldsymbol{Var(Y_i) = v(\mu_i)/\emptyset}$$

擬似尤度

分散$\boldsymbol{Var(Y_i)}$を扱うのはやめてある重み$\boldsymbol{V_i}$として考えてみる.

$$\boldsymbol{V_i = v(\mu_i)/\emptyset}$$

重み$\boldsymbol{V_i}$を用いた**擬似尤度スコア推定方程式**を解いて得られる推定量$\hat{\boldsymbol{\beta}}$ が，確率分布（分散）を仮定して計算した場合の最尤推定量$\boldsymbol{\beta}^*$と同じ性質（**一致性・漸近正規性**）を持てば妥当な統計的推論であるということ.

memo

Y_iの従う確率分布が指数分布族であるとき平均と分散は,

$$E(y_i) = \mu_i = a'(\theta_i),\quad Var(Y_i) = a''(\theta_i)/\emptyset$$

で，任意の関数 $v(\cdot)$ によって関係は一意に決まる.

$$Var(Y_i) = v(\mu_i)/\emptyset$$

擬似尤度スコア推定方程式

$$\boldsymbol{U_{QL}(\beta)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{K}} \boldsymbol{D_i}^{\boldsymbol{T}} \boldsymbol{V_i}^{-\boldsymbol{1}} \boldsymbol{S_i} = \boldsymbol{0}$$

where

$$D_i = \frac{\partial\{a'_i(\theta_i)\}}{\partial\beta} = A_i\Delta_i X_i$$

$$V_i = v(\mu_i)/\emptyset$$

$$S_i = Y_i - a'(\theta_i)$$

$$A_i = diag\{v(\mu_i)\}$$

memo

スコア推定方程式　$\boldsymbol{U_I(\beta)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{K}} \boldsymbol{X_i}^{\boldsymbol{T}} \Delta_i \boldsymbol{S_i} = \boldsymbol{0}$　*where*　$\Delta_i = diag(d\theta_{it}/d\eta_{it})$, $S_i = Y_i - a'(\theta_{it})$

漸近正規性の証明

$\boldsymbol{U_{QL}(\hat{\beta})}$を真値$\boldsymbol{\beta}^*$のまわりでテーラー展開する.

$$U_{QL}(\hat{\beta}) - U_{QL}(\beta^*) = \left\{\frac{\partial}{\partial\beta} U_{QL}(\beta)|_{\beta=\beta^*}\right\}(\hat{\beta} - \beta^*) + O_p(1)$$

$$\frac{1}{\sqrt{K}} U_{QL}(\beta^*) = \left\{-\frac{1}{K}\frac{\partial}{\partial\beta} U_{QL}(\beta)|_{\beta=\beta^*}\right\}\sqrt{K}(\hat{\beta} - \beta^*) + O_p(1)$$

$$\therefore \quad \boldsymbol{\sqrt{K}(\hat{\beta} - \beta^*) = \left\{-\frac{1}{K}\frac{\partial}{\partial\beta} U_{QL}(\beta)|_{\beta=\beta^*}\right\}^{-1} \frac{1}{\sqrt{K}} U_{QL}(\beta^*) + O_p(1)}$$

$$= \quad \boldsymbol{H^{-1}} \quad * \quad \mathbf{B} \quad + \boldsymbol{O_p(1)}$$

where

$$H = \left\{-\frac{1}{K}\frac{\partial}{\partial\beta} U_{QL}(\beta)|_{\beta=\beta^*}\right\}, \quad B = \frac{1}{\sqrt{K}} U_{QL}(\beta^*)$$

漸近正規性の証明（続き）

$$\therefore \quad \sqrt{\boldsymbol{K}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = \left\{-\frac{\mathbf{1}}{\boldsymbol{K}}\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{U}_{\boldsymbol{QL}}(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\right\}^{-1} \frac{\mathbf{1}}{\sqrt{\boldsymbol{K}}} \boldsymbol{U}_{\boldsymbol{QL}}(\boldsymbol{\beta}^*) + \boldsymbol{O}_{\boldsymbol{p}}(\mathbf{1})$$

$$= \quad \boldsymbol{H}^{-1} \quad * \quad \mathbf{B} \quad + \boldsymbol{O}_{\boldsymbol{p}}(\mathbf{1})$$

$$H = -\frac{1}{K}\left\{\frac{\partial}{\partial \beta} U_{QL}(\beta)|_{\beta=\beta^*}\right\} = -\lim_{K\to\infty} \frac{\partial}{\partial \beta} \frac{1}{K}\left\{\sum_{i=1}^{K} \left(\frac{\partial \mu_i}{\partial \beta}\right)^T V_i^{-1}(Y_i - \mu_i)\right\} = O_p(1)$$

（∵ 台数の法則）

$$B = \frac{1}{\sqrt{K}} U_{QL}(\beta^*) = \lim_{K\to\infty} \frac{1}{\sqrt{K}} \sum_{i=1}^{K} \left(\frac{\partial \mu_i}{\partial \beta}\right)^T V_i^{-1}(Y_i - \mu_i) \quad \sim Normal\ distribution$$

（∵ 中心極限定理）

$$\therefore \quad \sqrt{\boldsymbol{K}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \quad \sim \boldsymbol{Normal\ distribution}$$

一致性の証明：平均

平均 $E\left[\sqrt{K}\left(\hat{\beta}-\beta^{*}\right)\right]=E\left[\mathrm{HB}+O_p(1)\right]=O_p(1)*E[B]+O_p(1)$

$$E[B]=E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{K}\left\{\left(\frac{\partial\mu_i}{\partial\beta}\right)^{T}V_i{}^{-1}(Y_i-\mu_i)\right\}\right]=\frac{1}{\sqrt{K}}\sum_{i=1}^{K}\left\{\left(\frac{\partial\mu_i}{\partial\beta}\right)^{T}V_i{}^{-1}E[Y_i-\mu_i]\right\}=0$$

$$\therefore\quad \boldsymbol{E}\left[\sqrt{\boldsymbol{K}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{*}\right)\right]=\mathbf{0}$$

一致性の証明：分散

分散
$$Var\left[\sqrt{K}\left(\hat{\beta}-\beta^*\right)\right] = E\left[\{\sqrt{K}\left(\hat{\beta}-\beta^*\right)\}\{\sqrt{K}\left(\hat{\beta}-\beta^*\right)\}^T\right]$$
$$= E\left[\boldsymbol{H}^{-1}\boldsymbol{B}\boldsymbol{B}^{\boldsymbol{T}}(\boldsymbol{H}^{-1})^{\boldsymbol{T}}+O_p(1)\right]$$
$$= \boldsymbol{H}^{-1}E[\boldsymbol{B}\boldsymbol{B}^{\boldsymbol{T}}](\boldsymbol{H}^{-1})^{\boldsymbol{T}}$$

$$\boldsymbol{H} = -\frac{\partial}{\partial\beta}\frac{1}{K}\left\{\sum_{i=1}^{K}\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}(Y_i-\mu_i)\right\}$$
$$\to -\frac{\partial}{\partial\beta}E[\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}(Y_i-\mu_i)]$$
$$= -E[\left(\frac{\partial^2\mu_i}{\partial\beta^T\partial\beta}\right)^T V_i^{-1}(Y_i-\mu_i) - \left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}\left(\frac{\partial}{\partial\beta}V_i^{-1}\right)V_i^{-1}(Y_i-\mu_i)$$
$$+\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}\{\frac{\partial}{\partial\beta}(Y_i-\mu_i)\}]$$
$$= \left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}\left(\frac{\partial\mu_i}{\partial\beta}\right)$$

$$\therefore\ \boldsymbol{H} = \lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}\left(\frac{\partial\mu_i}{\partial\beta}\right)$$

一致性の証明：分散（続き）

分散

$$Var\left[\sqrt{K}\left(\hat{\beta}-\beta^*\right)\right] = E\left[\{\sqrt{K}\left(\hat{\beta}-\beta^*\right)\}\{\sqrt{K}\left(\hat{\beta}-\beta^*\right)\}^T\right]$$

$$= E\left[\boldsymbol{H}^{-1}\boldsymbol{B}\boldsymbol{B}^T(\boldsymbol{H}^{-1})^T + O_p(1)\right]$$

$$= \boldsymbol{H}^{-1}E[\boldsymbol{B}\boldsymbol{B}^T](\boldsymbol{H}^{-1})^T$$

$$= \boldsymbol{I_0}^{-1}\boldsymbol{I_1}\boldsymbol{I_0}^{-1}$$

$$\therefore\ \sqrt{\boldsymbol{K}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\right) \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I_0}^{-1}\boldsymbol{I_1}\boldsymbol{I_0}^{-1})$$

where

$$\boldsymbol{I_0} = \lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\left\{\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}\left(\frac{\partial\mu_i}{\partial\beta}\right)\right\}$$

$$\boldsymbol{I_1} = \lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\left\{\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}\cdot \underline{Var(Y_i)}\cdot V_i^{-1}\left(\frac{\partial\mu_i}{\partial\beta}\right)\right\}$$

Y_iが指数分布族に従うときは

$V_i = Var(Y_i)$なので$I_0 = I_1$

$\therefore\ \sqrt{K}\left(\hat{\beta}-\beta^*\right) \sim N(0, H^{-1})$

$q.e.d.$

$$E[\mathbf{BB^T}] = \lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\left\{\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}E[(Y_i-\mu_i)(Y_i-\mu_i)^T]V_i^{-1}\left(\frac{\partial\mu_i}{\partial\beta}\right)\right\}$$

$$= \lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\left\{\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}\cdot Var(Y_i)\cdot V_i^{-1}\left(\frac{\partial\mu_i}{\partial\beta}\right)\right\}$$

擬似尤度スコア推定方程式まとめ

・従属変数の真の分布が未知な場合にも，**重み**$\boldsymbol{V_i}$と**平均**$\boldsymbol{\mu_i}$の関係（任意の関数 $v(\cdot)$）さえ適当に与えてやれば，妥当な統計的推論を行える．

・特に，真の分布が指数分布であった場合($V_i = Var(Y_i)$)の推定量は，一般化線形モデルの最尤推定量と一致する．

・推定方程式は，陽に解くことは難しいので，反復重み付き二乗法など数値計算を行えば解が求まる．

e.g. Newton-Raphson method

一般化推定方程式(GEE)の着想

(Zeger, S. L., Liang, K. Y. , 1986)

従属変数が独立でない場合にも**妥当な**統計的推論を行いたい.

e.g. 土地利用など繰り返し観測データ

GLMの仮定の再確認

仮定①：従属変数が指数分布族に従う.

仮定②：従属変数がそれぞれ独立である. 尤度関数 $L(\theta|y) = \prod_{i=1}^{K} f(y_i)$

仮定②も緩める：$Cov(Y_i)$を考える.

従属変数Y_iが独立でない（相関構造を持つ）確率分布を見つけるのは難しいので，仮定①の制約条件を緩めた**擬似尤度スコア推定方程式の考え方を応用**して，仮定②の制約条件を取り除くことを考える.

“working” correlation matrix

$$\boldsymbol{V_i} = \boldsymbol{A}^{1/2}\boldsymbol{R_i(\alpha)}\boldsymbol{A}^{1/2}/\emptyset$$

“working” correlation matrix

where

$$\boldsymbol{A_i} = diag\{v(\boldsymbol{\mu_i})\}$$

$\boldsymbol{R_i}$：相関行列であるという要件を満たすn×n対称行列

α： $\boldsymbol{R}$を完全に特徴付けるs×1ベクトル

$\emptyset$：スケールパラメータ

擬似尤度スコア推定方程式　$U_{QL}(\beta) = \sum_{i=1}^{K} D_i{}^T V_i{}^{-1} S_i = 0$

相関構造を表現できるようにする
[パラメーターαを導入]
↓

where

$$D_i = \frac{\partial\{a'_i(\theta_i)\}}{\partial\beta} = A_i\Delta_i X_i$$

$$S_i = Y_i - a'(\theta_i)$$

$$V_i = v(\mu_i)/\emptyset$$

$$A_i = diag\{v(\mu_i)\}$$

一般化推定方程式：GEE

$$\boldsymbol{U_{GEE}(\beta,\alpha)} = \sum_{i=1}^{K} \boldsymbol{D_i}^{T}\boldsymbol{V_i}^{-1}\boldsymbol{S_i} = \boldsymbol{0}$$

where

$$\boldsymbol{D}_i = \frac{\partial\{a'_i(\theta_i)\}}{\partial\boldsymbol{\beta}} = \frac{\partial\mu_i}{\partial\boldsymbol{\beta}} = A_i\Delta_i X_i$$

$$\boldsymbol{V}_i = \boldsymbol{A}^{1/2}\boldsymbol{R_i(\alpha)}\boldsymbol{A}^{1/2}/\emptyset$$

$$\boldsymbol{S_i} = \boldsymbol{Y_i} - \boldsymbol{\mu_i(\beta)} = \boldsymbol{Y_i} - \boldsymbol{a'(\theta_i)}$$

$$\boldsymbol{A}_i = diag\{v(\boldsymbol{\mu}_i)\}$$

擬似尤度スコア推定方程式と同様．パラメーターαに依存することのみ異なる．

GEEの仮定と証明

(i) $\sqrt{K}(\hat{\alpha}-\alpha)=O_p(1)$ $\quad$ *given* β *and* ϕ

(ii) $\sqrt{K}\left(\hat{\phi}-\phi\right)=O_p(1)$ $\quad$ *given* β

(iii) $|\partial\hat{\alpha}(\beta,\phi)/\partial\phi| \le H(Y,\beta)$ $\quad$ *which is* $O_p(1)$

$U_{GEE}(\beta,\alpha)=\sum_{i=1}^{K} D_i{}^T V_i{}^{-1} S_i=0$ は次のように書き直せる.

α, β, φ依存

$$\sum_{i=1}^{K} \boldsymbol{U}_i\left\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}\{\boldsymbol{\beta},\hat{\boldsymbol{\phi}}(\boldsymbol{\beta})\}\right\}=\boldsymbol{0}$$

また，便宜上 $\boldsymbol{\alpha}^*(\boldsymbol{\beta})=\hat{\boldsymbol{\alpha}}\{\boldsymbol{\beta},\hat{\phi}(\boldsymbol{\beta})\}$ を定義しておく.

αの漸近正規性の証明

$\boldsymbol{\alpha}^*(\boldsymbol{\beta}) = \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\}$を$\boldsymbol{\alpha}$のまわりでテーラー展開する.

$$\frac{\sum_{i=1}^{K} U_i\{\beta, \alpha^*(\beta)\}}{\sqrt{K}} = \frac{\sum_{i=1}^{K} U_i(\beta, \alpha)}{\sqrt{K}} + \frac{\sum_{i=1}^{K} \partial U_i(\beta, \alpha)/\partial \alpha}{\sqrt{K}} (\alpha^* - \alpha) + O_p(1)$$

$$= \frac{\sum_{i=1}^{K} U_i(\beta, \alpha)}{\sqrt{K}} + \frac{\sum_{i=1}^{K} \partial U_i(\beta, \alpha)/\partial \alpha}{K} \sqrt{K}(\alpha^* - \alpha) + O_p(1)$$

$$= \boldsymbol{P}^* + \boldsymbol{Q}^*\boldsymbol{R}^* + O_p(1)$$

where

$$\boldsymbol{P}^* = \frac{\sum_{i=1}^{K} U_i(\beta, \alpha)}{\sqrt{K}} \qquad \boldsymbol{Q}^* = \frac{1}{K} \lim_{K\to\infty} \sum_{i=1}^{K} \boldsymbol{D_i}^{\boldsymbol{T}} \boldsymbol{V_i}^{-\boldsymbol{1}} \boldsymbol{S_i} = O_p(1)$$

$$\boldsymbol{R}^* = \sqrt{K}\left(\widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\} - \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \phi\} + \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \phi\} - \boldsymbol{\alpha}\right)$$

$$= \sqrt{K}\left\{(\hat{\phi} - \phi)\frac{\partial \widehat{\boldsymbol{\alpha}}}{\partial \phi}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\} + \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \phi\} - \boldsymbol{\alpha}\right\} = O_p(1)$$

$$\frac{\partial \widehat{\boldsymbol{\alpha}}}{\partial \phi}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\} = \lim_{\hat{\phi}\to\phi} \frac{\widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\} - \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \phi\}}{\hat{\phi} - \phi}$$

αの漸近正規性の証明（続き）

$\boldsymbol{\alpha}^*(\boldsymbol{\beta}) = \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\}$を$\boldsymbol{\alpha}$のまわりでテーラー展開する.

$$\frac{\sum_{i=1}^{K} U_i\{\beta, \alpha^*(\beta)\}}{\sqrt{K}} = \frac{\sum_{i=1}^{K} U_i(\beta, \alpha)}{\sqrt{K}} + \frac{\sum_{i=1}^{K} \partial U_i(\beta, \alpha)/\partial \alpha}{\sqrt{K}}(\alpha^* - \alpha) + O_p(1)$$

↑
漸近的に$\boldsymbol{P}^*$に一致

$$= \frac{\sum_{i=1}^{K} U_i(\beta, \alpha)}{\sqrt{K}} + \frac{\sum_{i=1}^{K} \partial U_i(\beta, \alpha)/\partial \alpha}{K}\sqrt{K}(\alpha^* - \alpha) + O_p(1)$$

$$= \boldsymbol{P}^* + \boldsymbol{Q}^*\boldsymbol{R}^* + O_p(1)$$

where

$$\boldsymbol{P}^* = \frac{\sum_{i=1}^{K} U_i(\beta, \alpha)}{\sqrt{K}} \qquad \boldsymbol{Q}^* = \frac{1}{K}\lim_{K\to\infty}\sum_{i=1}^{K} \boldsymbol{D_i}^{\boldsymbol{T}}\boldsymbol{V_i}^{-\boldsymbol{1}}\boldsymbol{S_i} = O_p(1)$$

↑
漸近的に正規分布

$$\boldsymbol{R}^* = \sqrt{K}\left(\widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\} - \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \phi\} + \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \phi\} - \boldsymbol{\alpha}\right)$$

$$= \sqrt{K}\left\{(\hat{\phi} - \phi)\frac{\partial \widehat{\boldsymbol{\alpha}}}{\partial \phi}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\} + \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \phi\} - \boldsymbol{\alpha}\right\} = O_p(1)$$

αの一致性の証明：平均と分散

$\frac{\sum_{i=1}^{K} U_i\{\beta, \alpha^*(\beta)\}}{\sqrt{K}}$の**平均**と**分散**は，擬似尤度スコア推定方程式の場合に導出したのと同様に導出出来る.

$$E\left[\frac{\sum_{i=1}^{K} U_i\{\boldsymbol{\beta}, \alpha^*(\boldsymbol{\beta})\}}{\sqrt{K}}\right] = \mathbf{0}$$

$$Var\left[\frac{\sum_{i=1}^{K} U_i\{\boldsymbol{\beta}, \alpha^*(\boldsymbol{\beta})\}}{\sqrt{K}}\right] = \lim_{K\to\infty}\{\sum_{i=1}^{K} {D_i}^T {V_i}^{-1} \cdot Cov(Y_i) \cdot {V_i}^{-1} {D_i}^T\}$$

βの漸近正規性の証明

$\boldsymbol{U_{GEE}}(\widehat{\boldsymbol{\beta}})$を真値$\boldsymbol{\beta}^*$のまわりでテーラー展開することで，擬似尤度スコア推定方程式の場合と同様に議論できて，

$$\sqrt{\boldsymbol{K}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)=\left\{-\frac{\mathbf{1}}{\boldsymbol{K}}\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{U_{GEE}}(\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta}))|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\right\}^{-\mathbf{1}}\frac{\mathbf{1}}{\sqrt{\boldsymbol{K}}}\boldsymbol{U_{GEE}}\left(\boldsymbol{\beta},\boldsymbol{\alpha}^*(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\right)+\boldsymbol{O_p}(\mathbf{1})$$

$$\therefore\quad \sqrt{\boldsymbol{K}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)\sim\boldsymbol{N}(\mathbf{0},\boldsymbol{J_0}^{-\mathbf{1}}\boldsymbol{J_1}\boldsymbol{J_0}^{-\mathbf{1}})$$

$where$

$$\boldsymbol{J_0}=\lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\left\{\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}\left(\frac{\partial\mu_i}{\partial\beta}\right)\right\}$$

$$\boldsymbol{J_1}=\lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\left\{\left(\frac{\partial\mu_i}{\partial\beta}\right)^T V_i^{-1}\cdot Cov(Y_i)\cdot V_i^{-1}\left(\frac{\partial\mu_i}{\partial\beta}\right)\right\}$$

$q.e.d.$

一般化推定方程式の数値計算

一般化推定方程式 $\sum_{i=1}^{K} \boldsymbol{U}_i \left\{\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \widehat{\boldsymbol{\phi}}(\boldsymbol{\beta})\}\right\} = \boldsymbol{0}$ は陽に解くことはできないので数値計算で解く.

$\widehat{\boldsymbol{\beta}}$と$\widehat{\boldsymbol{\alpha}}$の二段階推定

Modified Fisher scoring algorithm

$$\widehat{\boldsymbol{\beta}}_{j+1} = \widehat{\boldsymbol{\beta}}_j - \left\{\sum_{i=1}^{K} \boldsymbol{D}_i^{\boldsymbol{T}}(\widehat{\boldsymbol{\beta}}_j) \cdot \widetilde{\boldsymbol{V}_i}^{-1}(\widehat{\boldsymbol{\beta}}_j) \cdot \boldsymbol{D}_j(\widehat{\boldsymbol{\beta}}_j)\right\}^{-1} \cdot \left\{\sum_{i=1}^{K} \boldsymbol{D}_i^{\boldsymbol{T}}(\widehat{\boldsymbol{\beta}}_j) \cdot \widetilde{\boldsymbol{V}_i}^{-1}(\widehat{\boldsymbol{\beta}}_j) \cdot \boldsymbol{S}_i(\widehat{\boldsymbol{\beta}}_j)\right\}$$

$$where \quad \widetilde{\boldsymbol{V}_i}^{-1}(\boldsymbol{\beta}) = \boldsymbol{V}_i[\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \widehat{\boldsymbol{\phi}}(\boldsymbol{\beta})\}]$$

$$define \quad D = \left(D_1^{T}, \dots, D_K^{T}\right)^{T}, S = \left(S_1^{T}, \dots, S_K^{T}\right)$$

$$\widetilde{V}_i : nK \times nK \ block\ diagonal\ matrix$$

$$the\ modified\ dependent\ variable \quad \boldsymbol{Z} = \boldsymbol{D\beta} - \boldsymbol{S}$$

一般化推定方程式まとめ

- (回帰係数でなく分散に関心がある場合には使えないが) **従属変数が独立でない場合**にも，**分散の仮定なし**で妥当な統計的推論を行えるように，“working” correlation matrixを導入することで一般化線形モデルを拡張したものである.

参考文献

・ Shannon, C. E., A mathematical theory of communication, Part I, Part II. Bell Syst. Tech. J., Vol.27, pp.623-656, 1948.

・Wedderburn, R. W., Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method. Biometrika, Vol.61, No.3, pp.439-447, 1974.

・Zeger, S. L., Liang, K. Y., Longitudinal data analysis for discrete and continuous outcomes. Biometrics, pp.121-130, 1986.

・Liang, K. Y., Zeger, S. L., Longitudinal data analysis using generalized linear models. Biometrika, Vol.73, No.1, pp.13-22, 1986.

・柳本武美, 推定方程式に基づく推定 最尤法とモーメント法から. 応用統計学, Vol.24, No.1, pp.1-12, 1995.

・池田思朗, 情報源の構造推定: 確率モデルの構造探索アルゴリズム, Doctoral Dissertation, 1996.

・甘利俊一(2014). 情報幾何学の新展開. サイエンス社.

・Sopasakis, A. et al., Information metrics for improved traffic model fidelity through sensitivity analysis and data assimilation. Transportation Research Part B: Methodological, Vol.86, pp.1-18, 2016.