論　文

正規線形内包モデルとVuong検定

松　尾　精　彦*

要　約

正規線形モデルの一般化として，従来さまざまなモデルやモデル族が紹介され，その一般化に対応する推測理論が提案されてきている．たとえば，正規分布の仮定の代わりに分散均一性のみを仮定するガウス・マルコフモデル，平均パラメータのモデル化に非線形関数を許容する非線形正規モデル，正規分布から指数型分布族への拡張と平均パラメータ構造にリンク関数を導入した一般化線形モデル，さらに，一般化線形モデルの分布族の代わりに分布の平均分散関係のみを指定する擬似尤度一般化線形モデルなどの一般化が行われてきている．

中でも一般化線形モデルは，それまで別々に扱われてきていた重要ないくつかのモデルに，共通の枠組みを与えた点で大いに評価されるモデル族であると言える．この共通の枠組みの存在は，共通の推測理論や推定アルゴリズムの存在を意味していて，そのことが一般化線形モデルの存在意義を高めている．

しかし，この一般化線形モデルを，Vuong 検定というモデル選択のための手法を当てはめる立場から見直してみると，必ずしも扱いやすいモデルであるとは言いがたい面がある．Vuong 検定の枠組みでは，一般化線形モデルの長所が有効に働かないし，一般化線形モデルの仮定が不自然に見えてしまう場面が存在する．

そこで注目されるのが，正規線形モデルを内包するモデル族である．具体的には，プロビットモデルや対数線形正規モデル等がある．これらのモデルは，一般化線形モデルに属しているものもあるし，そうでないものもあるが，多くの場合，競合する一般化線形モデルに属するモデルと判別が難しいと言われている．この論文では，上記の正規線形内包モデルを，一般化線形モデルと比較しながら紹介し，Vuong 検定の枠組みでの優位性について議論する．

キーワード：Vuong 検定；一般化線形モデル；正規線形モデル
経済学文献季報分類番号：16-10

1　はじめに

正規線形モデルとは, n 個の確率変数, $Y_1, Y_2, \cdots, Y_n$, が独立に分散一定の正規分布に従っていて, それぞれの平均パラメータ $\mu_i := E[Y_i]$ が既知説明変数ベクトル $\boldsymbol{z}_i$ と未知パラメータベクトル $\boldsymbol{\beta}$ により $\mu_i = \boldsymbol{z}_i'\boldsymbol{\beta}$ と表される場合を言う．正規線形モデルでは, 推定量や検定統計量が正

*関西大学経済学部教授　amatsuo@ipcku.kansai-u.ac.jp

確な分布に従っていることから, 最も扱いやすい回帰モデルであり, 他の回帰モデルは正規線形モデルを可能な限り自然に拡張・一般化したものとして捉えることができる.

このような正規線形モデルの拡張・一般化の例として挙げられるのが, 分布の仮定を "無相関で分散が一定の無名の分布" に置き換えたガウス・マルコフモデル, 平均パラメータの線形性の制約を取り除いた非線形正規モデル, 分布の仮定を指数型分布族に拡げ平均パラメータの線形構造にリンク関数を導入した一般化線形モデル, さらに指数型分布族の代わりに分布の平均分散関係のみを仮定する擬似尤度一般化線形モデル等である.

この中でも, 最も興味深いのが Nelder and Wedderburn (1972) の一般化線形モデル (Generalized Linear Models) である. このモデルは, それ以前から重要なモデルとして用いられてきた, 二項分布回帰モデルであるロジットモデル, ポアソン分布回帰モデルである対数線形モデルをはじめ, ガンマ分布回帰モデル, そして当然ながら正規線形モデルが共有する性質により定義されたものである. これらのモデルの共通部分とは, 分布が指数型分布族であることと, 平均パラメータ構造が線形予測子 (linear predictor), $z_i'\boldsymbol{\beta}$, を内包することである. これらの共通部分により, 一般化線形モデルは共通の推定アルゴリズムを持つことになり, 尤度理論が統計的推測の漸近的性質を保障することになる.

正規線形モデルに対するガウス・マルコフモデルがあるように, 一般化線形モデルに対しては擬似尤度が導入されている. これは, Lindsey (1996) も主張するように, 実際のデータ生成メカニズムが, 分布の関数形を明確に指定するモデル内に含まれることはありえないという常識的な考え方にたいする対応と考えてよい. モデルはあくまでも現実を単純化したものに過ぎないという認識から, 分布の指定を平均分散関係というゆるやかなものにすることで, 現実に近づけようとする試みと言える.

一方, Vuong 検定は, 真の分布は候補に挙がっているモデルに含まれていても含まれていなくてもよいという前提のもとで, 分布の関数を明確に指定するモデル間の優劣を検定するものである. これは, White (1982) 他の misspecified models の議論を展開したものと言える. この Vuong 検定を実行する際に障害になるのは, 共通部分を有する2つのモデル間では, 尤度比検定統計量が重み付きカイ自乗和分布に従うということであり, 正確な棄却限界値が得られないという事実である. 一般化線形モデルでは, 特殊なモデルを除いては, この障害を避けて通れない.

そこで注目されるのが, 正規線型モデルの構造を含むモデル群の存在である. 具体的にはプロビットモデル, 対数正規モデルを挙げることができる. このようなモデルでは, 一定の追加的な条件下で, 正規線形モデルが持つ情報量行列が等値であるという性質を共有する. このことは, Vuong 検定における重み付きカイ自乗和分布が, 重み付けの無い普通のカイ自乗分布になることを意味する. それゆえ, Vuong 検定を行う際の障害が取り払われてしまうのである.

この論文では, 正規線形モデルを内包するモデル群を提案し, その性質を一般化線形モデルと

の比較を通して明らかにしてゆく．第 2 節では Vuong 検定を紹介し，第 3 節では Vuong 検定の利用を意識した一般化線形モデルの紹介を行う，その上で第 4 節では正規線形内包モデルを提案してその良好な性質について議論する．

2　Vuong Test

次のような状況を考えよう．確率変数のペア $(\boldsymbol{Y}, \boldsymbol{Z})$, $\boldsymbol{Y}$: l 次元確率変数, $\boldsymbol{Z}$: k 次元確率変数, がある．ここで, $\boldsymbol{Z} = \boldsymbol{z}$ が与えられたときの $\boldsymbol{Y}$ の条件付分布についての推測を行う場面を考える．

$\boldsymbol{X}_t = (\boldsymbol{Y}_t', \boldsymbol{Z}_t')'$ とおくとき, $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_t, \cdots, \boldsymbol{X}_n$ は互いに独立で観測可能な $m = l + k$ 次元確率変数列とする．ここで, $\boldsymbol{Y}_t'$ は $\boldsymbol{Y}_t$ の転置を表すものとする．このとき, 真の条件付分布を, $h^0(\boldsymbol{y}|\boldsymbol{z})$, $\boldsymbol{Z}$ の真の分布を $h^0(\boldsymbol{z})$, 2 つの競合する条件付モデルを,

$$(1) \qquad \boldsymbol{F_\theta} \equiv \left\{ f(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta}) \; ; \; \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathcal{R}^p \right\}, \quad \boldsymbol{G_\phi} \equiv \left\{ g(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\phi}) \; ; \; \boldsymbol{\phi} \in \boldsymbol{\Phi} \subset \mathcal{R}^q \right\}$$

と表すことにしよう．

Vuong 検定の枠組みでは, モデルが真の分布を含んでいても, いなくてもよいという立場をとるため, パラメータの pseudo-true value という考え方を導入する必要がある．$E^0_{\boldsymbol{X}}$ を $\boldsymbol{X}$ の真の分布 $h^0(\boldsymbol{x}) := h^0(\boldsymbol{y}|\boldsymbol{z})h^0(\boldsymbol{z})$ に関する期待値を表すものとするとき, パラメータの pseudo-true value は次のように与えられる．

$$(2) \qquad \boldsymbol{\theta}_* = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} E^0_{\boldsymbol{X}}[\ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta})], \quad \boldsymbol{\phi}_* = \arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} E^0_{\boldsymbol{X}}[\ln g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi})]$$

これは分布間の距離を Kullback-Leibler 情報量で測定するとき, 真の分布, $h^0(\boldsymbol{y}|\boldsymbol{z})$, との距離,

$$I\big(h^0(\boldsymbol{y}|\boldsymbol{z}), f(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta})\big) = E^0_{\boldsymbol{X}}[\ln h^0(\boldsymbol{Y}|\boldsymbol{Z})] - E^0_{\boldsymbol{X}}[\ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta})]$$

を最小にするモデル内の分布に対応する．

$\hat{\boldsymbol{\theta}}_n$, $\hat{\boldsymbol{\phi}}_n$ を, それぞれ, モデル $F_{\boldsymbol{\theta}}$, $G_{\boldsymbol{\phi}}$ における最尤推定量とする．

$$\hat{\boldsymbol{\theta}}_n = arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_n^f(\boldsymbol{\theta}), \quad \text{where } L_n^f(\boldsymbol{\theta}) \equiv \sum_{t=1}^{n} \ln f(\boldsymbol{y}_t|\boldsymbol{z}_t;\boldsymbol{\theta}),$$

$$\hat{\boldsymbol{\phi}}_n = arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} L_n^g(\boldsymbol{\phi}), \quad \text{where } L_n^g(\boldsymbol{\phi}) \equiv \sum_{t=1}^{n} \ln g(\boldsymbol{y}_t|\boldsymbol{z}_t;\boldsymbol{\phi}).$$

すると, 適当な正則条件の下で, $\hat{\boldsymbol{\theta}}_n$ と $\hat{\boldsymbol{\phi}}_n$ は, それぞれ $\boldsymbol{\theta}_*$ と $\boldsymbol{\phi}_*$ の強一致推定量である．また, $\sqrt{n}\,(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)$ と $\sqrt{n}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_*)$ は, それぞれ, 漸近的に正規分布

$$(3) \qquad N\Big[\boldsymbol{0}, A_f^{-1}(\boldsymbol{\theta}_*)B_f(\boldsymbol{\theta}_*)A_f^{-1}(\boldsymbol{\theta}_*)\Big], \quad N\Big[\boldsymbol{0}, A_g^{-1}(\boldsymbol{\phi}_*)B_g(\boldsymbol{\phi}_*)A_g^{-1}(\boldsymbol{\phi}_*)\Big]$$

に従う．ここで，

$$\begin{cases} A_f(\boldsymbol{\theta}) \equiv E^0_{\boldsymbol{X}}[\dfrac{\partial^2 \ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}], \\ B_f(\boldsymbol{\theta}) \equiv E^0_{\boldsymbol{X}}[\dfrac{\partial \ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\dfrac{\partial \ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}]. \end{cases}$$

$\hat{\boldsymbol{\theta}}_n$ と $\hat{\boldsymbol{\phi}}_n$ の同時分布もまた漸近正規であり，

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_* \\ \hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_*\end{pmatrix} \sim N\Big[\begin{pmatrix}\mathbf{0} \\ \mathbf{0}\end{pmatrix}, \Sigma_{(\boldsymbol{\theta}_*,\boldsymbol{\phi}_*)}\Big], \tag{4}$$

ここで，

$$\Sigma_{(\boldsymbol{\theta}_*,\boldsymbol{\phi}_*)} = \begin{pmatrix} A_f^{-1}(\boldsymbol{\theta}_*)B_f(\boldsymbol{\theta}_*)A_f^{-1}(\boldsymbol{\theta}_*) & A_f^{-1}(\boldsymbol{\theta}_*)B_{fg}(\boldsymbol{\theta}_*,\boldsymbol{\phi}_*)A_g^{-1}(\boldsymbol{\phi}_*) \\ A_g^{-1}(\boldsymbol{\phi}_*)B_{gf}(\boldsymbol{\theta}_*,\boldsymbol{\phi}_*)A_f^{-1}(\boldsymbol{\theta}_*) & A_g^{-1}(\boldsymbol{\phi}_*)B_g(\boldsymbol{\phi}_*)A_g^{-1}(\boldsymbol{\phi}_*) \end{pmatrix}$$

ただし，

$$B_{fg}(\boldsymbol{\theta},\boldsymbol{\phi}) \equiv E^0_{\boldsymbol{X}}[\frac{\partial \ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial \ln g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}].$$

“どちらのモデルがより真の分布に近い分布を含むか” という問題は，次のような仮説に言い換えられる．帰無仮説，

$$H_0^V : E^0_{\boldsymbol{X}}[\ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)] = E^0_{\boldsymbol{X}}[\ln g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)], \tag{5}$$

にたいし，

$$H_f^V : E^0_{\boldsymbol{X}}[\ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)] > E^0_{\boldsymbol{X}}[\ln g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)], \tag{6}$$

あるいは，

$$H_g^V : E^0_{\boldsymbol{X}}[\ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)] < E^0_{\boldsymbol{X}}[\ln g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)], \tag{7}$$

を対立仮説とする検定を行う．

Remark 1 H_0^V が棄却され H_f^V が採択される場合，

$$I\big(h^0(\boldsymbol{y}|\boldsymbol{z}), g(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\phi}_*)\big) > I\big(h^0(\boldsymbol{y}|\boldsymbol{z}), f(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta}_*)\big) \geq 0$$

という関係が成り立ち G_ϕ は真の分布を含み得ない．また，H_0^V の下では，$f(\cdot|\cdot;\boldsymbol{\theta}_*) \neq g(\cdot|\cdot;\boldsymbol{\phi}_*)$ ならば，2 つのモデルはどちらも真の分布を含まないことに注意されたい．

検定統計量として，対数尤度比，

$$LR_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\phi}}_n) \equiv L_n^f(\hat{\boldsymbol{\theta}}_n) - L_n^g(\hat{\boldsymbol{\phi}}_n) = \sum_{t=1}^n \ln \frac{f(\mathbf{Y}_t|\boldsymbol{Z}_t;\hat{\boldsymbol{\theta}}_n)}{g(\mathbf{Y}_t|\boldsymbol{Z}_t;\hat{\boldsymbol{\phi}}_n)} \tag{8}$$

を用いる．尤度関数の適当な条件下で，大数の法則より，

$$\frac{1}{n}LR_n(\hat{\boldsymbol{\theta}}_n,\hat{\boldsymbol{\phi}}_n) \to E^0_{\boldsymbol{X}}\left[\ln\frac{f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)}{g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)}\right] \text{ a.s.}, \tag{9}$$

が成立することが今後用いられる．また，$\ln\dfrac{f(\boldsymbol{Y}_t|\boldsymbol{Z}_t;\boldsymbol{\theta}_*)}{g(\boldsymbol{Y}_t|\boldsymbol{Z}_t;\boldsymbol{\phi}_*)}$ の分散を，

$$\begin{aligned}\omega^2_* &\equiv \mathrm{var}^0_{\boldsymbol{X}}\left[\ln\frac{f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)}{g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)}\right]\\ &= E^0_{\boldsymbol{X}}\left[\ln\frac{f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)}{g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)}\right]^2 - \left[E^0_{\boldsymbol{X}}\left[\ln\frac{f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)}{g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)}\right]\right]^2\end{aligned} \tag{10}$$

で表すことにする．一般の尤度理論と違い，モデルが真の分布を含んでいることを前提としないことから情報量等式 $A_f(\boldsymbol{\theta}) \equiv -B_f(\boldsymbol{\theta})$ が成り立つ保証がないので，次の重み付きカイ自乗和分布を導入する必要が生じる．

Definition 1 (重み付きカイ 2 乗和分布): $Z_1, Z_2, \cdots, Z_m$ を独立で標準正規分布に従う確率変数列とし，$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_m)$ を m 個の実数からなるベクトルとする．このとき，確率変数 $\sum_{i=1}^m \lambda_i Z_i^2$ の従う分布を，重み付きカイ 2 乗和分布と呼びその累積分布関数を $M_m(\cdot, \boldsymbol{\lambda})$ で表す．

Theorem 1 (尤度比検定統計量の漸近分布): 適当な正則条件下で，

(i) $f(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta}_*) = g(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\phi}_*)$ のとき，

$$2LR_n(\hat{\boldsymbol{\theta}}_n,\hat{\boldsymbol{\phi}}_n) \to M_{p+q}(\cdot;\boldsymbol{\lambda}_*) \text{ in law},$$

が成り立つ．ここで $\boldsymbol{\lambda}_*$ は，

$$W = \begin{bmatrix} -B_f(\boldsymbol{\theta}_*)A_f^{-1}(\boldsymbol{\theta}_*) & -B_{fg}(\boldsymbol{\theta}_*,\boldsymbol{\phi}_*)A_g^{-1}(\boldsymbol{\phi}_*) \\ B_{gf}(\boldsymbol{\phi}_*,\boldsymbol{\theta}_*)A_f^{-1}(\boldsymbol{\theta}_*) & B_g(\boldsymbol{\phi}_*)A_g^{-1}(\boldsymbol{\phi}_*) \end{bmatrix},$$

の固有値からなるベクトルである．

(ii) $f(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta}_*) \neq g(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\phi}_*)$ のとき，

$$n^{-1/2}LR_n(\hat{\boldsymbol{\theta}}_n,\hat{\boldsymbol{\phi}}_n) - n^{1/2}E^0_{\boldsymbol{X}}\left[\ln\frac{f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)}{g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)}\right] \to N(0,\omega^2_*) \text{ in law},$$

が成り立つ．

この定理の大まかな証明をしよう．pseudo true value における尤度関数を，最尤推定値の周りで Taylor 展開して，

$$L^f_n(\boldsymbol{\theta}_*) = L^f_n(\hat{\boldsymbol{\theta}}_n) + \frac{n}{2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)'A_f(\boldsymbol{\theta}_*)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) + o_p(1).$$

$$L_n^g(\phi_*) = L_n^g(\hat{\phi}_n) + \frac{n}{2}(\hat{\phi}_n - \phi_*)' A_g(\phi_*)(\hat{\phi}_n - \phi_*) + o_p(1).$$

を得る．これら 2 式の差より，

$$\begin{aligned} LR_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\phi}}_n) =& LR_n(\boldsymbol{\theta}_*, \boldsymbol{\phi}_*) - \frac{n}{2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*)' A_f(\boldsymbol{\theta}_*)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \\ & + \frac{n}{2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_*)' A_g(\boldsymbol{\phi}_*)(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_*) + o_p(1), \end{aligned} \tag{11}$$

が導かれる．(i) の場合，$f(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta}_*) = g(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\phi}_*)$ だから，$LR_n(\boldsymbol{\theta}_*, \boldsymbol{\phi}_*) \equiv 0$ が成り立ち，

$$LR_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\phi}}_n) = \frac{n}{2}\begin{pmatrix} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_* \\ \hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_* \end{pmatrix}' \begin{bmatrix} -A_f & 0 \\ 0 & A_g \end{bmatrix} \begin{pmatrix} \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_* \\ \hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_* \end{pmatrix} + o_p(1)$$

となる．$(\hat{\boldsymbol{\theta}}_n \vdots \hat{\boldsymbol{\phi}}_n)$ の漸近分布は (4) であることより (i) が示される．

一方 (ii) は，(11) の両辺から $nE^0_{\boldsymbol{X}}[\ln \frac{f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)}{g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)}]$ を引き，$\sqrt{n}$ で割ることにより，

$$\begin{aligned} & n^{-1/2} LR_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\phi}}_n) - n^{1/2} E^0_{\boldsymbol{X}}[\ln \frac{f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)}{g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)}] \\ & \qquad = n^{1/2}\Big[\frac{1}{n} LR_n(\boldsymbol{\theta}_*, \boldsymbol{\phi}_*) - E^0_{\boldsymbol{X}}[\ln \frac{f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)}{g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)}]\Big] + o_p(1). \end{aligned} \tag{12}$$

を得るが，右辺は $LR_n(\boldsymbol{\theta}_*, \boldsymbol{\phi}_*)$ が退化しないことから，中心極限定理を適用することにより，$N(0, \omega_*^2)$ に漸近的に従うことから示される．

この定理は，Vuong (1989) が紹介したモデル選択検定の枠組みの中で，中心的な役割を果たすものであり，(i) については適用上困難さが付きまとうが，(ii) はどのようなモデルに対しても簡単に適用できるという利点を持つ．

2.1　Strictly Non-nested Models

Definition 2 (Strictly Non-Nested Models): 2 つの条件付モデル $F_{\boldsymbol{\theta}}$ と $G_{\boldsymbol{\phi}}$ は，

$$F_{\boldsymbol{\theta}} \cap G_{\boldsymbol{\phi}} = \emptyset.$$

であるとき strictly non-nested であると言う．

Example 1 (Cox, 1961): 対数正規分布と指数分布，

$$\begin{cases} f(y|\theta) = y^{-1}(2\pi\theta_2)^{-1/2} \exp\Big\{-\dfrac{(\ln y - \theta_1)^2}{2\theta_2}\Big\},\ 0 < \theta_2 < \infty,\ y > 0. \\ g(y|\phi) = \phi^{-1}\exp(-y/\phi),\ \phi > 0,\ y > 0. \end{cases}$$

をそれぞれ条件付分布とするモデルを考える．なお，対数正規分布の場合は θ_1 に，指数分布の場合には，ϕ に線形構造を入れる．

上の例の他にも, 対数正規分布とワイブル分布, また, 対数正規分布とガンマ分布についても競合する分布として比較・議論がされてきている.

Example 2 probit モデルと logit モデル: 説明変数が与えられたときの, 二項変量 Y の確率が累積正規分布関数で表されるのが probit モデルで, ロジスティック分布の累積分布関数で表されるのが logit モデルである.

$$\begin{cases} \Pr(Y=1) = \Phi(\boldsymbol{\theta}'\boldsymbol{z}) = \displaystyle\int_{-\infty}^{\boldsymbol{\theta}'\boldsymbol{z}} \frac{1}{\sqrt{2\pi}} \exp\Big\{-\frac{1}{2}u^2\Big\}du, \\ \Pr(Y=1) = \Delta(\boldsymbol{\phi}'\boldsymbol{z}) = \dfrac{e^{\boldsymbol{\phi}'\boldsymbol{z}}}{1+e^{\boldsymbol{\phi}'\boldsymbol{z}}}. \end{cases}$$

これら 2 つのモデルは, いずれも, 次の節で紹介する一般化線形モデルに属しているもので, リンク関数の異なるモデルと言える.

以上のように, strictly non-nested case の定義を厳密に適用するなら, 2 つのモデルの誤差分布が互いに異なる場合しかない. それゆえ, 2 つの正規線形回帰モデルの場合は, 特別の前提を置かない限り, strictly non-nested ではない.

さて strictly non-nested case では 2 つのモデルに共通な分布が存在しないため, $f(\cdot|\cdot;\boldsymbol{\theta}_*) \neq g(\cdot|\cdot;\boldsymbol{\phi}_*)$ であることが保証されるので, **Theorem 1** の (ii) を直接適用できる. この場合特に, 次の定理を得る.

Theorem 2 (Model Selection Tests for Strictly Non-Nested Models): 適当な正則条件の下で, $F_{\boldsymbol{\theta}}$ と $G_{\boldsymbol{\phi}}$ が strictly non-nested ならば, 次が成り立つ.

$$(i) \text{ under } H_0^{\mathrm{V}} : n^{-1/2}LR_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\phi}}_n)/\hat{\omega}_n \to N(0,1) \text{ in law },$$

$$(ii) \text{ under } H_f^{\mathrm{V}} : n^{-1/2}LR_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\phi}}_n)/\hat{\omega}_n \to +\infty \text{ a.s. },$$

$$(iii) \text{ under } H_g^{\mathrm{V}} : n^{-1/2}LR_n(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\phi}}_n)/\hat{\omega}_n \to -\infty \text{ a.s. }.$$

上の定理の ω_*^2 の推定量として, Vuong (1989) が列挙している漸近同値なものの中で, 例えば,

$$\hat{\omega}_n^2 \equiv \frac{1}{n}\sum_{t=1}^{n}\Big[\ln\frac{f(\boldsymbol{y}_t|\boldsymbol{z}_t;\hat{\boldsymbol{\theta}}_n)}{g(\boldsymbol{y}_t|\boldsymbol{z}_t;\hat{\boldsymbol{\phi}}_n)}\Big]^2 - \Big[\frac{1}{n}\sum_{t=1}^{n}\ln\frac{f(\boldsymbol{y}_t|\boldsymbol{z}_t;\hat{\boldsymbol{\theta}}_n)}{g(\boldsymbol{y}_t|\boldsymbol{z}_t;\hat{\boldsymbol{\phi}}_n)}\Big]^2$$

を用いればよいだろう.

2.2 Overlapping Models

Definition 3 (Overlapping Models): 2 つの条件付モデル $F_{\boldsymbol{\theta}}$ と $G_{\boldsymbol{\phi}}$ が overlapping であるとは,

$$(i)\ F_{\boldsymbol{\theta}} \cap G_{\boldsymbol{\phi}} \neq \emptyset,$$

$$(ii)\ F_{\boldsymbol{\theta}} \setminus G_{\boldsymbol{\phi}} \neq \emptyset \text{ and } G_{\boldsymbol{\phi}} \setminus F_{\boldsymbol{\theta}} \neq \emptyset.$$

であるときを言う.

Example 3 2つの線形回帰モデルで,

$$\begin{cases} f(y|\boldsymbol{z};\boldsymbol{\theta}) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\Big\{-\dfrac{(y-\alpha-\boldsymbol{u}'\boldsymbol{\beta}-\boldsymbol{v}_f'\gamma_f)^2}{2\sigma^2}\Big\}, \\ g(y|\boldsymbol{z};\boldsymbol{\phi}) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\Big\{-\dfrac{(y-\alpha-\boldsymbol{u}'\boldsymbol{\beta}-\boldsymbol{v}_g'\gamma_g)^2}{2\sigma^2}\Big\}, \end{cases}$$

ここで $\boldsymbol{u}, \boldsymbol{v}_f, \boldsymbol{v}_g$ は, それぞれ p, q_f, q_g 次元確率変数ベクトルで, $(\boldsymbol{u}\vdots\boldsymbol{v}_f\vdots\boldsymbol{v}_g)$ は退化しない $p+q_f+q_g$ 次元確率変数ベクトルであるとする. ここで, $(\boldsymbol{u}\vdots\boldsymbol{v}_f\vdots\boldsymbol{v}_g)$ はベクトルを縦に並べたもの, つまり $(\boldsymbol{u}', \boldsymbol{v}_f', \boldsymbol{v}_g')'$ を表すものとする. ここでは正規線形回帰モデルを考えたが, 他にも, 一般化線形モデルを始めとする回帰モデル一般に広く例を求めることが出来る.

Example 4 Gamma 分布回帰モデルと Weibull 分布回帰モデル: 指数分布回帰モデルを共有する.

この例に限らず, あるモデルの2通りの拡張が考えられる場合には overlapping case となる. 一般化線形モデルで言えば, 分布に2通りの拡張法がある場合や, link 関数に2つの拡張法がある場合がそうである.

この場合は, strictly non-nested case よりも扱いが厄介になる. なぜなら2つのモデルに共通部分があるため, モデル選択検定を行う際には, 先ず **Theorem 1** の *(i)* を適用して, $f(\cdot|\cdot;\boldsymbol{\theta}_*) = g(\cdot|\cdot;\boldsymbol{\phi}_*)$ であるか否かを確かめなければならない. $f(\cdot|\cdot;\boldsymbol{\theta}_*) = g(\cdot|\cdot;\boldsymbol{\phi}_*)$ と判定されれば, 帰無仮説 H_0^{V} が採択される. そうでなければ, 引き続き $f(\cdot|\cdot;\boldsymbol{\theta}_*) \neq g(\cdot|\cdot;\boldsymbol{\phi}_*)$ のもとで, strictly non-nested の場合と同様, **Theorem 1** の *(ii)* を適用するのである.

Remark 2 どちらかのモデルが真の分布を含むという古典的な仮定を置くならば, $f(\cdot|\cdot;\boldsymbol{\theta}_*)$ あるいは $g(\cdot|\cdot;\boldsymbol{\phi}_*)$ が真の分布 $h^0(\cdot|\cdot)$ に一致するので, $E^0_{\boldsymbol{X}}[\ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)] = E^0_{\boldsymbol{X}}[\ln g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)]$ は $I\big(h(\boldsymbol{y}|\boldsymbol{z}), f(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta}_*)\big) = I\big(h(\boldsymbol{y}|\boldsymbol{z}), g(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\phi}_*)\big) = 0$ を意味し, Kullback-Leibler 情報量の性質より, $h(\boldsymbol{y}|\boldsymbol{z}) = f(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta}_*) = g(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\phi}_*)$ という関係が得られる. つまり,

$$f(\cdot|\cdot;\boldsymbol{\theta}_*) = g(\cdot|\cdot;\boldsymbol{\phi}_*) \Leftrightarrow E^0_{\boldsymbol{X}}[\ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)] = E^0_{\boldsymbol{X}}[\ln g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)]$$

が成立する. それ故 $f(\cdot|\cdot;\boldsymbol{\theta}_*) \neq g(\cdot|\cdot;\boldsymbol{\phi}_*)$ ならば $E^0_{\boldsymbol{X}}[\ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)] \neq E^0_{\boldsymbol{X}}[\ln g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)]$ であることが導かれるので, **Theorem 1** の *(ii)* を適用する必要が無くなり, $f(\cdot|\cdot;\boldsymbol{\theta}_*) = g(\cdot|\cdot;\boldsymbol{\phi}_*)$ を帰無仮説とする検定を行いさえすればよいことになる.

2.3 Nested Models

Definition 4 (Nested Models): 条件付モデル G_ϕ が条件付モデル F_θ に nest in しているとは,

$$G_\phi \subset F_\theta.$$

であるときを言う.

Example 5 2 つの正規線形回帰モデルで, 片方のモデルが採用している説明変数が他方のそれの部分集合であるとき.

$$\begin{cases} f(y|\boldsymbol{z};\boldsymbol{\theta}) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\dfrac{(y-\alpha-\boldsymbol{u}'\boldsymbol{\beta}-\boldsymbol{v}'\boldsymbol{\gamma})^2}{2\sigma^2}\Big\}, \\ g(y|\boldsymbol{z};\boldsymbol{\phi}) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\dfrac{(y-\alpha-\boldsymbol{u}'\boldsymbol{\beta})^2}{2\sigma^2}\Big\}, \end{cases}$$

ここで $\boldsymbol{u}$, $\boldsymbol{v}$ は, それぞれ p, q 次元確率変数ベクトルで, $(\boldsymbol{u}\vdots\boldsymbol{v}) = (\boldsymbol{u}', \boldsymbol{v}')'$ は退化しない $p+q$ 次元確率変数ベクトルであるとする. この場合も, 先の overlapping case と同様, 回帰モデル全般に例を求めることが出来る.

Example 6 指数分布回帰モデルと Gamma(Weibull) 分布回帰モデルのように, 一方のモデルが他方のモデルを拡張したものである場合.

Remark 3 モデルの包含関係より $E^0_{\boldsymbol{X}}[\ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)] \geq E^0_{\boldsymbol{X}}[\ln g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)]$ が成り立つため, $H^{\mathbb{V}}_g$ は必然的に除外される. また, $E^0_{\boldsymbol{X}}[\ln f(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\theta}_*)] = E^0_{\boldsymbol{X}}[\ln g(\boldsymbol{Y}|\boldsymbol{Z};\boldsymbol{\phi}_*)]$ は $f(\cdot|\cdot;\boldsymbol{\theta}_*) = g(\cdot|\cdot;\boldsymbol{\phi}_*)$ を意味するので, **Theorem 1** の *(i)* を適用して, $f(\cdot|\cdot;\boldsymbol{\theta}_*) = g(\cdot|\cdot;\boldsymbol{\phi}_*)$ を検定しさえすればよいことが分かる.

3 Models

この節では, 前節で紹介した Vuong 検定を実際に適用するモデルの具体例を挙げて行くことにする. まず最初は正規線形モデルから開始し, 次に一般化線形モデルを紹介する. その上で, 一般化線形モデルを更に拡張したモデルや, 以前から競合するモデルとして用いられてきたモデルを紹介し, Vuong 検定の適用例を紹介してゆく.

なお, 今後は, Y_t はスカラー値確率変数, $\boldsymbol{Z}_t$ は k 次元確率変数とし, $(Y_t, \boldsymbol{Z}_t)$, $t = 1, 2, \cdots, n$ は独立で同一な分布に従う n 個の確率変数列とする.

3.1　正規線形モデル

$\boldsymbol{Z}=\boldsymbol{z}$ が与えられたときの Y の条件付分布が, 以下の密度関数をもつと仮定するのが, 正規線形モデルである.

$$f(y|\boldsymbol{z};\boldsymbol{\theta})=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y-\boldsymbol{z}'\boldsymbol{\beta})^2}{2\sigma^2}\right\}$$

この正規線形モデルを仮定するときの扱いやすさは, ここで改めて説明する必要が無いほどよく知られている. 例えば, 佐和 (1979) を参考にされたい. 平均パラメータが未知パラメータの線形結合として表せること, 誤差分布が独立で分散が均一な正規分布に従うことの 2 点が, 数学的な扱いやすさを生じさせている.

しかし, 見方を変えれば, この 2 つは, 過剰な制約ではないかとも言える. これに対し正規線形モデルが, 見かけよりも, 広範な適応範囲を持つことを次のように説明できる. まず正規分布の仮定だが, その正当性は中心極限定理に求めることができる. 中心極限定理にはさまざまな形式があるが (例えば, 清水 (1976) を参照されたい), その一つを証明抜きで与えよう.

Theorem 3 独立な確率変数列 $\{E_i\}$, $i=1,2,3,\cdots$ の平均を μ_i 分散を σ_i^2 とし,

$$\lim_{n\to\infty}\frac{\max_{1\le j\le n}\sigma_j^2}{\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2}=0$$

が成り立つとき, あと少しの仮定を加えたら,

$$\frac{(E_1-\mu_1)+(E_2-\mu_2)+\cdots+(E_n-\mu_n)}{\sqrt{\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2}}$$

は漸近的に正規分布に従う.

上の定理を利用して, 正規線形モデルの一つの解釈を与えよう. 被説明変数 Y はさまざまな変動要素の無限和として,

$$Y=E_1'+E_2'+\cdots+E_n'+\cdots$$

と表されているとしよう. この段階では, $\{E_j'\}$, $j=1,2,3,\cdots$ は中心極限定理の要件を満たしている必要は無い.

この状況で, 有力な説明変量として $\{Z_j\}$, $j=1,2,\cdots,p$ という確率変数を導入し, 上の $\{E_j'\}$, $j=1,2,3,\cdots$ から $\{Z_j\}$, $j=1,2,\cdots,p$ の影響を取り除いたものを, $\{E_j\}$, $j=1,2,3,\cdots$ とし,

$$Y=\beta_0+Z_1\beta_1+Z_2\beta_2+\cdots+Z_p\beta_p+E_1+E_2+\cdots$$

と変形する, ここで, $\{E_j\}$, $j=1,2,3,\cdots$ が中心極限定理の要件を満たすならば,

$$\epsilon=E_1+E_2+E_3+\cdots$$

は正規分布 $\mathrm{N}(0,\sigma^2)$ に従うと考えてよい. つまり, $Z_1, Z_2, \cdots, Z_p$ で条件付けた Y の分布が正規分布となる.

この解釈をさらに進めよう. 上に述べた $\{Z_j\}$, $j=1,2,\cdots,p$ の選び方は一通りではなく, かなりの数の組み合わせが考えられる. すると, 条件付ける確率変数を変えれば, それに対応した真の分布が与えられる. Vuong 検定においては, $\boldsymbol{Z}: k \times 1$ random vector をあらかじめ与えて議論しているため, 真の分布は一意に決定される. しかし, $\boldsymbol{Z}: k \times 1$ random vector 以外の確率変数ベクトルで条件付けを行っても, 正規線形モデルが得られる可能性がある. nested models としてこのようなもの (パラメータ空間に制約を加えたモデルと解釈できる) が選ばれたとすれば, 松尾 (2004) にもあるように, $-A_f(\boldsymbol{\theta}_*) = B_f(\boldsymbol{\theta}_*), -A_g(\boldsymbol{\phi}_*) = B_g(\boldsymbol{\phi}_*)$, が成立する. なぜなら, 説明変数で条件付けたときの条件付分布が正しい分布であるからである. それゆえ χ^2 (重み付きではない) 統計量を用いた Nested models の検定が可能になる.

3.2 一般化線形モデル

この節では, 正規線形モデルの拡張である一般化線形モデルを, Vuong 検定の枠組みを意識しながら紹介する. 分布の関数形を明確に規定する一般化線形モデルは, 正規線形モデルの分布の指定を正規分布から指数型分布族に拡張し, 平均パラメータ構造にリンク関数を許容することで平均構造を拡張したものといえる. 正規線形モデルに対するガウス・マルコフモデルと同様に, 一般化線形モデルには誤差分布の平均分散関係のみを仮定するものもあるが, 関数形が指定されなければ Vuong 検定は実行できないので, ここでは触れない.

まず, 分布は正規分布から, 以下のような構造を持つ分布族に拡張する（分布の拡張).

(13) $$f_Y(y;\theta,\phi) = \exp\Big\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\Big\}$$

この分布族は, θ を自然パラメータとする一母数指数型分布族に, 散らばり母数 ϕ を付け加えたものである. この ϕ を追加することにより, 正規分布, ガンマ分布, 逆ガウス分布が含まれるようになる. 上の密度関数より, 平均パラメータ μ と自然パラメータ θ との関係や分散が以下のように求められる.

(14) $$E[Y] = \mu = b'(\theta), \quad V[Y] = a(\phi)b''(\theta)$$

次に平均パラメータの線形構造を次のように一般化する（リンク関数 g の導入).

(15) $$g(\mu) = \eta = \boldsymbol{z}'\boldsymbol{\beta}$$

このように一般化することで, 線形構造 $\eta := \boldsymbol{z}'\boldsymbol{\beta}$ を保ったままの拡張に制約している.

第 2 節で述べた例では全て線形構造 $\eta := \boldsymbol{z}'\boldsymbol{\beta}$ を持っているし, ほとんどの例では分布の仮定 (13) を満たしている. 例外はワイブル分布と対数正規分布を仮定するモデルのみであり, それ

以外は全て一般化線形モデルに属している．このように，一般化線形モデルは Vuong 検定を適用する対象を数多く含むモデルと言える．

なお，一般化線形モデルでは，散らばりパラメータ ϕ は平均パラメータの推定に影響を及ぼさないような形で付け加えられていて，$\boldsymbol{\beta}$ は最尤法により推定されるが，ϕ は局外パラメータ扱いでモーメント法で推定される．これは，正規線形モデルにおける分散パラメータと同じ扱いである．変数選択の際は，この局外パラメータの推定値を利用して，F 検定に似た手法が用いられる．これらの手続きは，もちろん，仮定されたモデルが正しい条件付き分布を含んでいるという古典的な前提のもとで正当化される．詳しくは，McCullagh and Nelder (1989) を参照されたい．

4　正規線形内包モデル

前々節の説明から分かるように, Vuong 検定は strictly non-nested models では利用しやすいが, overlapping models や nested models では, 帰無仮説 $f(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta}_*) = g(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\phi}_*)$ を検定しなければならないという厄介な問題を抱えている．この問題を軽減するようなモデル族を提案するのがこの分節の目的である．なお，overlapping models は 二組の nested models を合わせたものと解釈できる．それゆえ, 次のような nested models,

$$\boldsymbol{F}_\theta \equiv \left\{f(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\theta})\ \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathcal{R}^p\right\} \supset \boldsymbol{G}_\phi \equiv \left\{g(\boldsymbol{y}|\boldsymbol{z};\boldsymbol{\phi})\ \boldsymbol{\phi} \in \boldsymbol{\Phi} \subset \mathcal{R}^q\right\},$$

のみを対象にする．

3.1 で述べたように, 正規線形モデルでは, 条件付ける確率変数の組を取り替えることにより, 正しい条件付分布を含むモデルが複数得られる可能性がある．このことは, 正規分布の再生性 (X と Y が正規分布に従うとき, $X+Y$ もまた正規分布に従う) や，正規分布の平均パラメータが正規分布に従うときの複合分布もまた正規分布であるという事実を用いることにより説明できる．このように, 説明変数で条件付けたときの条件付分布が正規分布に従っている場合には, いくつかの正しい条件付分布を含むモデルが得られることになる．

このような性質は，正規分布しか持ち得ないもので, 他の分布で望むことはできない．一般化線形モデルの定義は, 尤度理論を用いる際には都合のよいものではあるが, 説明変数と誤差の関係が希薄であり, 天下り的なモデルと言うことができる．一般化線形モデルは, データ生成過程を確率論的に描写するものではなく, むしろ, データ生成過程を確率論的に描写するモデルを近似する, 数学的に扱いやすいモデルと言えるのではないか．このことを例を挙げながら説明してゆこう．

Example 7 (probit model): ϵ_i は正規分布 $N(0,\sigma^2)$ に従う確率変数とするとき,

$$Y_i' = \beta_0 + Z_{1i}\beta_1 + Z_{2i}\beta_2 + \cdots + Z_{pi}\beta_p + \epsilon_i,$$

と表されるとしよう. Y_i' の値がある限界値 λ を超えたら $Y_i = 1$, そうでなければ $Y_i = 0$ となる場合を考えよう. すると Y_i は成功確率,

$$\Pr\{Y_i = 1|\boldsymbol{z}\} = 1 - \Phi(\frac{\lambda - \boldsymbol{z}'\boldsymbol{\beta}}{\sigma}) = \Phi(\frac{-\lambda + \boldsymbol{z}'\boldsymbol{\beta}}{\sigma}) \tag{16}$$

のベルヌイ試行となり, probit model に帰着する. 競合するモデルは logit model である. logit model は，十分統計量が存在するという特長をもち理論的に扱い易いのだが, データ生成過程を説明するものではない. これら probit model と logit model とは，お互いを判別するのが事実上不可能なくらい, よく似たモデルであることは良く知られている (竹内・藤野，1980).

Example 8 (lognormal model): Y_i' は対数正規分布に従う確率変数, $\log Y' \sim N(0,\sigma^2)$, であるとき,

$$Y_i \sim e^{\beta_0 + Z_{1i}\beta_1 + Z_{2i}\beta_2 + \cdots + Z_{pi}\beta_p}\ Y_i',$$

と表されるとしよう. このとき, $\boldsymbol{Z}_i$ の実現値で条件付けたときの Y_i の分布は, 対数正規分布に従う. このモデルは, 正規線形モデルの性質を受け継いでいる. 競合するモデルは gamma 分布 $G(\alpha, 1/\alpha)$ を用いたものであり，一般化線形モデルに属する.

この例を見ても分かるように, 一般化線形モデルでは説明変数と誤差分布との関係が不自然であり, 数学的処理を優先させたモデルと言える. なお, この 2 つのモデルの有効性比較については, Firth (1988) を参考にされたい.

Example 9 (Poisson model): Y を稀現象の大量観察により得られる確率変数としよう. つまりポアソン分布に従うことが想定されるものとしよう. このとき Y の期待値 $\mu = \mathrm{E}(Y)$ は, 観察の期間 (あるいは規模) に比例すると仮定することは合理的である. このような場合,

$$\ln(\mu) = \ln(\text{time length}) + \ln(\text{other effects})$$

となる. 実際, McCullagh & Nelder(1989), p.206 では次のようなモデルを採用している.

$$\begin{aligned}
&\ln(\text{expected number of damage incidents}) \\
&\quad = \beta_0 + \ln(\text{aggregate months service}) \\
&\qquad\quad + \ln(\text{effect due to ship type}) \\
&\qquad\quad + \ln(\text{effect due to year of construction}) \\
&\qquad\quad + \ln(\text{effect due to service period})
\end{aligned} \tag{17}$$

このデータでは, 1. ship type, 2. year of construction, 3.service periood という 3 つの名義尺度を用いて損傷事故を分類し, 件数を合計している. aggregate month service とは, 該当する船の稼動期間を合計したものである. すると, i-th ship type, j-th year of construction and k-th service period に該当する l-th ship の事故数を $Y_{ijk,l}$ と置くとき, ポアソン分布にしたがい, その期待値は,

$$\ln \mu_{ijk,l} = \ln E(Y_{ijk,l}) = \beta_0 + \ln T_{ijk,l} + \ln \alpha_i + \ln \beta_j + \ln \gamma_k + \epsilon_{ijk,l}$$

と表してよい. 実際のデータは, $Y_{ijk} = \sum_l Y_{ijk,l}$ であり, 平均が,

$$\mu_{ijk} = \sum_l \mu_{ijk,l} = \sum_l e^{\alpha_i+\beta_j+\gamma_k} e^{T_{ijk,l}} e^{\epsilon_{ijk,l}}$$

のポアソン分布に従うとしてよい. $\epsilon_{ijk,l}$ を正規分布 $N(0,\sigma^2)$ とすると lognormal Poisson 分布モデルとなり, $e^{T_{ijk,l}}$ を $Gamma(\alpha, \frac{1}{\alpha})$ とすると Gamma Poisson 分布モデルとなる.

Gamma Poisson 分布モデルは, 一般化線形モデルの枠組みで議論されるモデルであるが, 前の例と同様に, lognormal Poisson 分布モデルを近似した数学的に扱いやすいモデルと言える.

上の 3 つの例を見て分かるように, 一般化線形モデルに属するモデルは, 正規線形モデルを内包するモデルと競合関係にある. ここで正規線形内包モデルを次のように定義しておこう.

Definition 5 正規線形内包モデルとは, i 番目の観測変量が, 次のような形の要素を内包するものである.

$$\beta_0 + Z_{1i}\beta_1 + Z_{2i}\beta_2 + \cdots + Z_{pi}\beta_p + \epsilon_i,$$

ここで, $Z_1, Z_2, \cdots, Z_p$ は説明変量で, ϵ_i は正規分布に従う誤差とする.

以上のモデルは, 一般化線形モデルに含まれているものもあればそうでないものもある. これらのモデルの利点として, 中心極限定理に基づくデータ生成プロセスとして自然なものであること, そして, 変数選択検定を行う際 χ^2 検定の妥当性が高いということが挙げられる. このモデルは正規線形モデルの拡張と言え, 別な方向への拡張としての一般化正規モデルに競合するものと考えてもよい.

この論文の終わりに, 正規線形内包モデルの持つ特長についてまとめよう.

1. 正規線形内包モデルでは, 正しい条件付モデルが複数存在する可能性がある. つまり, 正規線形内包モデルは, 正しい条件付モデルをピンポイントで持つものではなく, いくつかの正しい条件付モデルを許容する構造になっている. それゆえ各モデルにおいて情報量等値が成り立ち, 尤度比検定統計量が, 重み付きカイ自乗和分布ではなく, カイ自乗分布に従う.

2. 正規線形内包モデルは，データ生成過程を描写するが，一般化線形モデルは天下り的なモデルが多い．正規線形内包モデルとそれに競合する一般化線形モデルとを判別することは難しいと言われている．

5 終わりに

この論文では, 正規線形内包モデルを紹介し，その長所について述べた．それに競合する一般化線形モデルでは，利用可能な説明変数の数が限られている場面での利用が目立つ．これに対し，経済や会計の分野では説明変数は数多く存在し, さまざまな説明変数の組が取り上げられる場面が多い．こういった説明変数が数多く存在する場面では，誤差と説明変数の間には緊密な関係がある．ある説明変数を外すと, その説明変数は誤差として扱われるのである．それゆえ，正規線形内包モデルが，より自然なモデルとして想定されるのである．

参考文献

[1] Cox, D.R. (1961) Tests of separate families of hypotheses. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.*

[2] Firth, D. (1988) Multiplicative errors: log-normal or gamma? *J.R.Statist.Soc.* **B 50**, pp266-268.

[3] Lindsey J.K. (1996) *Parametric Statistical Inference*, Oxford University Press.

[4] Nelder, J.A. and Wedderburn R.W.M. (1972) Generalized linear models, *J.R.Statist.Soc.*, **A 135**, 370-84.

[5] Vuong, Q.H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, Vol. 57, 307-333.

[6] White H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica*, Vol. 50, 1-25.

[7] 稲垣 宣生 (2003) 数理統計学 改訂版, 裳華房.

[8] 佐和 隆光 (1979) 回帰分析, 朝倉書店.

[9] 清水 良一 (1976) 中心極限定理, 教育出版.

[10] 竹内 啓, 藤野 和建 (1980) 二項分布とポアソン分布, 東京大学出版会.

[11] 松尾 精彦 (2003) モデル選択基準とその正規線形回帰モデルへの適用. 関西大学経済論集, 第 53 巻, 93–107.

[12] 松尾 精彦 (2004) Vuong test とその正規線形回帰モデルへの適用法. 関西大学経済論集, 第 54 巻, 39–60.